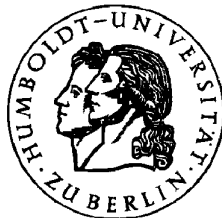


Einflussfaktoren auf Berliner Immobilienpreise Eine statistische Analyse

Diplomarbeit

zur Erlangung des Grades
einer Diplom-Kauffrau

an der Wirtschaftswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin



vorgelegt von

Edna Freese

(Matrikel-Nr. 138984/97)

1. Prüfer: Prof. Dr. Wolfgang Härdle

Berlin, 22. April 2003

Inhaltsverzeichnis

Abkürzungsverzeichnis	3
Tabellenverzeichnis	5
Abbildungsverzeichnis	7
1 Problemstellung und Vorgehensweise	8
2 Merkmale von Immobilienmärkten	10
3 Die Häuser	14
3.1 Der ursprüngliche Datensatz	14
3.2 Modifikationen des Datensatzes	16
4 Die Bezirke	19
4.1 Beschreibung der Bezirke	19
4.2 Kaufpreisbeeinflussende Bezirkseigenschaften	22
4.3 Zusammenfassung der Hypothesen	27
5 Statistische Grundlagen	29
5.1 Hedonische Regression	29
5.2 Das allgemeine lineare Regressionsmodell	30
5.2.1 Form der Funktion	30
5.2.2 Dummies und kategorielle Variablen	32
5.2.3 Annahmen	32
5.2.4 Parameterschätzung nach der Methode der kleinsten Quadrate	33
5.2.5 Eigenschaften des Kleinst-Quadrate-Schätzers	34
5.2.6 Parameterschätzung nach der Maximum-Likelihood-Me- thode	35
5.2.7 Das Bestimmtheitsmaß	36
5.2.8 Robuste Schätzungen	36
5.2.9 Interpretation der Koeffizienten	36

<i>INHALTSVERZEICHNIS</i>	2
6 Deskriptive Statistiken	39
6.1 Deskriptive Statistiken zu den Häusern	39
6.2 Deskriptive Statistiken zu den Bezirken	44
7 Missings	49
7.1 Struktur der missings	49
7.2 Das Mittlere monatliche Haushaltsnettoeinkommen	53
7.3 Arbeitslose	55
8 Regression	63
8.1 Vorgehensweise	63
8.2 Regression auf den Gesamtkaufpreis über den gesamten Zeitraum 1991 bis 2000	66
8.2.1 Residuen- und Ausreißeranalyse	66
8.2.2 Bezirksdummies	69
8.2.3 Regressionskennzahlen	71
8.2.4 Koeffizienten der Hauseigenschaften und deren Interpretation	71
8.2.5 Koeffizienten der Bezirkseigenschaften und deren Interpretation	74
8.2.6 Koeffizienten der Jahresdummies und deren Interpretation	78
8.3 Regression auf den Gesamtkaufpreis über den Zeitraum 1996 bis 2000	79
8.3.1 Vorgehensweise und Ausreißeranalyse	79
8.3.2 Regressionskennzahlen für die Teilregression	82
8.3.3 Koeffizienten der Teilregression: Hauseigenschaften	83
8.3.4 Koeffizienten der Teilregression: Lageeigenschaften	86
8.3.5 Koeffizienten der Teilregression: Bezirkseigenschaften	87
8.3.6 Koeffizienten der Teilregression: Jahresdummies	88
8.3.7 Vergleich der Regressionen	89
9 Zusammenfassung und Ausblick	90
A Liste der Variablen	92
A.1 Die Hausvariablen und deren Erklärung	92
A.2 Die Lagevariablen und deren Erklärung	98
A.3 Die Bezirksvariablen und deren Erklärung	102
B Beobachtungen nach Bezirk und Jahr	108
Literaturverzeichnis	111

Abkürzungsverzeichnis

BauGB	Baugesetzbuch
BauNVO	Baunutzungsverordnung
BauO Bln	Bauordnung für Berlin
BGB	Bürgerliches Gesetzbuch
BLUE	Best Linear Unbiased Estimator
BMZ	Baummassenzahl
BUE	Best Unbiased Estimator
BVG	Berliner Verkehrsbetriebe
CBD	Central Business District
Charlottenb.	Charlottenburg
Coef.	Coefficient
durchschnittl.	durchschnittlich
GbR	Gesellschaft bürgerlichen Rechts
gemeinn.	gemeinnützig
GFZ	Geschossflächenzahl
ha	Hektar
HHNetto	Mittleres monatliches Haushaltsnettoeinkommen
Hohenschönh.	Hohenschönhausen
jur.	juristisch
km	Kilometer
KQ-Schätzer	Kleinste-Quadrate-Schätzer
Max	Maximum
Min	Minimum
ML	Maximum-Likelihood
MSE	Mean Square Error
Obs	Observation
öffentl.	öffentlich
OLS	Ordinary Least Squares
persönl.	persönlich
Prenzl. Berg	Prenzlauer Berg
SchulG	Schulgesetz für Berlin

sonst.	sonstige
Std. Dev.	Standard Deviation
tatsächl.	tatsächlich
ungewöhl.	ungewöhnlich
Verhältn.	Verhältnis
vs.	versus
WertV	Wertermittlungsverordnung
Wohnunt.	Wohnungsunternehmen

Tabellenverzeichnis

6.1	Deskriptive Statistiken für Alter, Grundfläche und Geschossfläche	40
6.2	Preise in den einzelnen Bezirken	40
6.3	Deskriptive Statistiken für den Preis, getrennt nach Ost und West	42
6.4	Anteil der Häuser an bestimmten Kategorien	43
6.5	Anteil der Häuser mit besonderen Eigenschaften	44
6.6	Deskriptive Statistiken für Berlin	44
6.7	Anteil verschiedener Teilflächen an der Gesamtfläche des Bezirks im Jahr 2000	45
6.8	Deskriptive Statistiken zu den Bezirken	46
7.1	Anzahl missings für Hausvariablen	50
7.2	Zuordnung der Häuser zu Lagekategorien	52
7.3	Deskriptive Statistiken für das Mittlere monatliche Haushaltsnettoeinkommen	54
7.4	Mittleres monatliches Haushaltsnettoeinkommen pro Bezirk	55
7.5	Arbeitslose im Jahresdurchschnitt	56
7.6	Abweichungsindex und MSE	59
7.7	Anzahl Arbeitslose im Jahresdurchschnitt 1991 - 2000	60
7.8	Mittlere Anzahl Arbeitslose sowie Anteil an Erwerbsbevölkerung und Gesamtbevölkerung pro Bezirk für den Zeitraum 1991 - 2000	61
7.9	Anteil Arbeitslose an Erwerbsbevölkerung und Gesamtbevölkerung für Gesamt-Berlin	62
8.1	Variablen x_k und Anzahl Beobachtungen, für die $DFBETA_{ki} > \frac{2}{\sqrt{n}}$	68
8.2	Regressionskennzahlen: Modell mit Bezirksdummies	69
8.3	Koeffizienten der Bezirksdummies	70
8.4	Regressionskennzahlen: Modell mit Bezirksvariablen	71

8.5	Koeffizienten der Hauseigenschaften	71
8.6	Koeffizienten der Bezirkseigenschaften	74
8.7	Koeffizienten der Jahresdummies	78
8.8	Anzahl missings für Lagevariablen	80
8.9	Variablen x_k und Anzahl Beobachtungen, für die $DFBETA_{ki} > \frac{2}{\sqrt{n}}$, für die Teilregression 1996 - 2000	80
8.10	Regressionskennzahlen der Teilregression: Modell mit Bezirksdummies	82
8.11	Regressionskennzahlen der Teilregression: Modell mit Bezirksvariablen	83
8.12	Regressionskennzahlen der Teilregression: Modell mit Bezirks- und Lagevariablen	83
8.13	Koeffizienten der Hauseigenschaften der Teilregression über den Zeitraum 1996 - 2000	84
8.14	Deskriptive Statistiken für Alter, Grundfläche, Geschossfläche und Preis für den Teilzeitraum 1996 - 2000	85
8.15	Koeffizienten der Lageeigenschaften der Teilregression über den Zeitraum 1996 - 2000	86
8.16	Koeffizienten der Bezirkseigenschaften der Teilregression über den Zeitraum 1996 - 2000	87
8.17	Koeffizienten der Jahresdummies der Teilregression über den Zeitraum 1996 - 2000	89
8.18	Korrigierte Bestimmtheitsmaße für die verschiedenen Regressionen	89
A.1	Liste der Hausvariablen	92
A.2	Liste der Lagevariablen	98
A.2	Obergrenzen für die Bestimmung des Maßes baulicher Nutzung nach §17 BauNVO	99
A.3	Liste der Bezirksvariablen	102
B.1	Relative und absolute Häufigkeiten der Beobachtungen nach Bezirk und Jahr	108

Abbildungsverzeichnis

4.1	Neue und alte Bezirke Berlins	22
6.1	Preisniveau in Berlin	41
6.2	Durchschnittliche Preise in Ost-, West- und Gesamt-Berlin über die Jahre 1991 - 2000	42
6.3	Anzahl der S- und U-Bahnstationen in den Bezirken im Jahr 2000	48
7.1	Das Mittlere monatliche Haushaltsnettoeinkommen in Zehlen- dorf mit linearer Regressionsgerade	54
7.2	Mittlerer Anteil Arbeitslose an der Bevölkerung und Standard- abweichung der Bezirke	62
8.1	$\ln(\text{Preis})$ vs. $\ln(\text{Grundstücksfläche})$, $\ln(\text{Preis})$ vs. $\ln(\text{Geschoss-}$ $\text{fläche})$, $\ln(\text{Preis})$ vs. Alter	64
8.2	Residual vs. Fitted Plot für den Gesamtzeitraum 1991 - 2000	69
8.3	Preisentwicklung in den Jahren 1991 bis 2000	79
8.4	Residual vs. Fitted Plot für den Teilzeitraum 1996 - 2000 . .	82

Kapitel 1

Problemstellung und Vorgehensweise

Beim Kauf eines Einfamilienhauses beeinflusst neben der Ausstattung des Objektes auch dessen Lage, bestimmt durch die Region oder den Bezirk, in der oder dem das Haus liegt, die Kaufentscheidung. Zwei Häuser mit derselben Ausstattung kosten je nach Bezirk unterschiedlich viel. Grund dafür ist die unterschiedliche Attraktivität der Bezirke. Gedanklich wird der Preis für ein bebautes Grundstück in einen Hauswert, der durch die Hauscharakteristika, und einen Bodenwert, der durch die Bezirkscharakteristika bestimmt wird, getrennt. In dieser Arbeit soll herausgefunden werden, welche Eigenschaften einen Bezirk besonders attraktiv machen, so dass für sie ein Preisaufschlag gezahlt wird, und welche weniger attraktiv sind, so dass sie nur für einen Preisabschlag in Kauf genommen werden. Da die Lagecharakteristika selbst nicht gehandelt werden und so ihre Preise nicht direkt beobachtet und bestimmt werden können, müssen diese indirekt aus den Preisen für ein bebautes Grundstück bestimmt werden. Eine Methode, indirekt Preise zu berechnen, ist die hedonische Methode. Danach besteht jedes Gut aus einem Bündel von Charakteristika, für die ein impliziter oder hedonischer Preis berechnet werden kann (vgl. Rosen (1974)). Übertragen auf diese Arbeit bedeutet das, dass ein bebautes Grundstück als ein Bündel von Haus- und Bezirkscharakteristika gesehen wird, für die ein impliziter Preis bestimmt werden soll. Das Hauptaugenmerk liegt dabei auf der Identifizierung von Bezirkseigenschaften, die die Attraktivität eines Bezirks und damit den Bodenwert bestimmen. Ziel ist die Ermittlung einer Preisfunktion mit Hilfe der hedonischen Methode. Dafür wird der Markt für Einfamilienhäuser in Berlin im Zeitraum 1991 bis 2000 betrachtet.

Die Vorgehensweise ist folgendermaßen: Im zweiten Kapitel werden Beson-

derheiten des Immobilienmarktes erläutert und gesetzliche Methoden der Wertermittlung dargestellt. Im dritten Kapitel wird zuerst der Datensatz vorgestellt sowie dessen Modifikationen erläutert. Daraufhin werden im vierten Kapitel die Berliner Bezirke kurz beschrieben. Es wird überlegt, welche Bezirkseigenschaften die Kaufentscheidung beeinflussen könnten. Es folgt im fünften Kapitel eine Erklärung der verwendeten Methode und im sechsten Kapitel eine ausführliche Beschreibung der Häuser und der Bezirke. Im siebten Kapitel wird auf missings eingegangen, bevor im achten Kapitel letztendlich mit der Analyse begonnen wird. Es werden zwei Regressionsfunktionen ermittelt: eine über den Gesamtzeitraum 1991 bis 2000, die andere über einen Teilzeitraum 1996 bis 2000. Der Gesamtkaufpreis wird als Summe aus Bauwert, ermittelt durch hedonische Regression auf Hauseigenschaften, und Bodenwert, ermittelt durch hedonische Regression auf Bezirkseigenschaften, betrachtet. In der zweiten Regression werden zusätzliche Variablen, die vorher nicht erfasst wurden und die Lage des Objekts genauer beschreiben, benutzt. Im letzten Kapitel erfolgt die Zusammenfassung und der Ausblick.

Kapitel 2

Merkmale von Immobilienmärkten

In diesem Kapitel sollen Besonderheiten des Immobilienmarktes im Vergleich zum vollkommenen Markt der ökonomischen Theorie erläutert und gesetzliche Vorschriften der Wertermittlung vorgestellt werden.

In der ökonomischen Theorie führt der Markt Angebot und Nachfrage zusammen. Sind die gehandelten Güter homogen, handelt es sich um einen vollkommenen Markt. Die Theorie der Preisbildung beruht auf folgenden Annahmen: Die Anbieter maximieren den Gewinn, die Nachfrager ihren Nutzen. Es herrscht vollkommene Konkurrenz durch eine große Anzahl an Marktteilnehmern (Polypol). Die Marktteilnehmer betrachten den Preis als gegeben und treffen ihre Entscheidungen unabhängig voneinander. Der Markt ist transparent, d. h. die für die Preisbildung wesentlichen Daten wie Güterqualität, Preise usw. sind bekannt, und der Marktzugang ist offen. Unter diesen Bedingungen bildet sich ein Gleichgewichtspreis, der Angebot und Nachfrage zum Ausgleich bringt (vgl. Woll (2000)).

Analog werden auf dem Immobilienmarkt Nachfrage und Angebot von Immobilien zusammengeführt. Im Gegensatz zum perfekten Markt der ökonomischen Theorie ist der Immobilienmarkt allerdings nicht perfekt. Wie Heuer (1985) ausführt, weisen sowohl das gehandelte Gut als auch der Markt an sich Besonderheiten auf. So sind die gehandelten Güter heterogen. Sie unterscheiden sich in Größe, Alter, Zustand, Lage usw. Sie sind nicht beweglich, sondern immobil. "Die Standortgebundenheit [...] bedingt, daß der Nutzen [...] u. a. von externen Faktoren determiniert wird. Solche externen Faktoren sind: die Entfernung zu Arbeitsstandorten, Einkaufsmöglichkeiten oder anderen infrastrukturellen Einrichtungen [...] und] die Qualität des Wohnumfeldes

[...]” (Heuer; 1985). Weitere Besonderheiten sind die Unteilbarkeit, die Dauerhaftigkeit, die lange Produktionsdauer, die die Anpassung an unterschiedliche Marktlagen erschwert, und die hohen Produktionskosten, durch die ”eine starke Abhängigkeit von den Verhältnissen am Kapitalmarkt” (Heuer; 1985) entsteht.

Der Wohnungsmarkt zerfällt durch die Immobilität des gehandelten Gutes in regionale Teilmärkte sowie durch die Heterogenität in sachliche Teilmärkte. Dieses Problem bleibt auch dann bestehen, wenn man sich, wie in dieser Arbeit, regional auf Berlin und sachlich auf Einfamilienhäuser beschränkt, da sich der Teilmarkt Berlin in weitere ”Unterteilmärkte”, die Bezirke, mit jeweils ihnen eigenen Charakteristika aufspaltet und auch Einfamilienhäuser sich in ihren Qualitätsmerkmalen deutlich unterscheiden. Diese Vielfalt von Angebot und Nachfrage führt zu einer mangelnden Markttransparenz und damit zu Schwierigkeiten bei der Preisfindung.

Der korrekten Bewertung von bebauten Grundstücken kommt insofern besondere Bedeutung zu, als das Immobilienvermögen den größten Aktivposten in der Vermögensbilanz privater Haushalte darstellt. ”Bewertet zu Wiederbeschaffungspreisen und abzüglich der Abschreibungen betrug es Ende 1997 schätzungsweise gut 7 Billionen DM” (Deutsche Bundesbank; 1999). Das entspricht 51% des Bruttovermögens. Die Eigentümerquote war 1998 in Berlin mit 11 % deutlich geringer als die Gesamtdeutsche Eigentümerquote von 40,9 %. Sie betrug im Westteil 13,2 %, im Ostteil 7,3 % (vgl. Statistisches Bundesamt (2000)). Die Ermittlung des Verkehrswerts ist in der Wertermittlungsverordnung (WertV) gesetzlich geregelt.

Der Sachwert bebauter Grundstücke setzt sich additiv aus dem Grundstücks-wert (Bodenwert) und dem Gebäudewert (Bauwert) zusammen.

§7 WertV kennt drei Verfahren zur Ermittlung des Verkehrswerts: das Vergleichswertverfahren, das Ertragswertverfahren und das Sachwertverfahren.

Der Bauwert wird durch das Ertragswertverfahren oder das Sachwertverfahren bestimmt (§§15, 21 WertV). Allerdings ”richtet sich [der Verkehrswert von Einfamilienhäusern] in der Regel nicht nach Renditeüberlegungen. Der Sachwert und die jeweilige Marktverfassung (Angebot und Nachfrage) sind die wertbestimmenden Elemente. [...] Reihenhäuser und andere standardisierte Typenhäuser, bei denen eine ausreichende Zahl von Vergleichsgrundstücken zur Verfügung steht, können im Vergleichswertverfahren bewertet werden” (Ross et al.; 1997). ”Bei Anwendung des Sachwertverfahrens ist der Wert der

baulichen Anlagen [...] getrennt vom Bodenwert nach Herstellungswerten zu ermitteln" (§21 Abs. 1 WertV). Somit umfasst der Bauwert "die gesamten Herstellungskosten eines Gebäudes zum Wertermittlungszeitpunkt" (Ross et al.; 1997). "Der Herstellungswert von Gebäuden ist unter Berücksichtigung ihres Alters [...] und von Baumängeln und Bauschäden [...] sowie sonstiger wertbeeinflussender Umstände [...] zu ermitteln" (§21 Abs. 3 WertV). "Nach der Ermittlung des Sachwertes kann es notwendig sein, noch weitere Wertkorrekturen vorzunehmen. [...] Das Problem [...] liegt in ihrer Nachvollziehbarkeit. [...] Daraus folgt, daß der Sachwert nicht unmittelbar als Grundlage des Verkehrswertes anzusehen ist" (Ross et al.; 1997).

Der Bodenwertanteil ist derjenige Anteil am Gesamtkaufpreis, der auf den Grund und Boden entfällt. Dieser "ist in der Regel im Vergleichswertverfahren [...] zu ermitteln" (§15 Abs. 2 WertV). "Bei der Anwendung des Vergleichswertverfahrens sind Kaufpreise solcher Grundstücke heranzuziehen, die hinsichtlich der ihren Wert beeinflussenden Merkmale [...] hinreichend übereinstimmen (Vergleichsgrundstücke)" (§13 Abs. 1 WertV). Ross et al. (1997) stellen dazu fest, dass sich das "Marktgeschehen auf dem Grundstücksmarkt [...] in einer Vielzahl von Einzelfällen mit ganz unterschiedlichen Grundstücken und Verhältnissen ab[spielt,] und nur selten findet man so weitgehende Übereinstimmung, daß die Wertverhältnisse eines Grundstücks unmittelbar für ein anderes als Bewertungsmaßstab übernommen werden können."

"Zur Ermittlung des Bodenwerts können neben oder anstelle von Preisen für Vergleichsgrundstücke auch geeignete Bodenrichtwerte herangezogen werden" (§13 Abs. 2 WertV). "Auf Grund der Kaufpreissammlung sind für jedes Gemeindegebiet durchschnittliche Lagewerte für den Boden unter Berücksichtigung des unterschiedlichen Entwicklungszustands [...] zu ermitteln (Bodenrichtwerte)" (§196 Abs. 1 S. 1 BauGB). "Zur Führung der Kaufpreissammlung ist jeder Vertrag [...] von der beurkundenden Stelle in Abschrift dem Gutachterausschuß zu übersenden" (§195 Abs. 1 S. 1 BauGB). "Die Bodenrichtwerte werden aus den vorliegenden Kaufpreisen abgeleitet" (Ross et al.; 1997). Die Kaufpreissammlung erhöht die Markttransparenz und reduziert so die Marktunvollkommenheit.

Nach §196 Abs. 1 BauGB sind Bodenrichtwerte in bebauten Gebieten "mit dem Wert zu ermitteln, der sich ergeben würde, wenn der Boden unbebaut wäre." Allerdings werden die "Kaufpreise [...] nur sehr selten in die auf den Grund und Boden und auf die Gebäude entfallenden Anteile unterteilt; so sind auch Bodenpreise, die von Gutachterausschüssen gesammelt und in Richtwertkarten zusammengefaßt werden, oft nur im Wege der Schätzung

oder in Fortschreibung früherer Werte entstanden und damit keine echten Vergleichspreise" (Ross et al.; 1997). Aufgrund der geschilderten Probleme bei der Ermittlung von Bau- und Bodenwert nach den gesetzlichen Verfahren wird in dieser Arbeit ein anderer Weg der Wertermittlung beschritten.

Eine weit verbreitete Möglichkeit zur Ermittlung von objektiven Preisen für heterogene Güter besteht in der Anwendung der hedonischen Methode. Der Grundgedanke ist, dass jedes Gut aus einem Bündel von Charakteristika besteht. Jedem Charakteristikum kann ein impliziter oder hedonischer Preis zugeordnet werden. Bezogen auf Immobilien bedeutet dies, dass implizite Preise für das Alter, den Zustand, die Blocklage usw. berechnet werden. Der Wert eines bebauten Grundstücks ergibt sich dann aus den Mengen an Charakteristika, die das Objekt aufweist, und deren impliziten Preisen (vgl. Rosen (1974)).

Da zwei Häuser gleicher Qualität dasselbe kosten sollten, sind Unterschiede im Kaufpreis auf unterschiedliche Bodenpreise zurückzuführen (z. B. Haurin und Brasington (1996)). Diese werden wiederum durch die Lage des Hauses bestimmt. Berlin bestand im Betrachtungszeitraum aus 23 sehr unterschiedlichen Bezirken, die anhand verschiedener Variablen beschrieben werden. Es wird davon ausgegangen, dass Hauseigenschaften wie Quadratmeterzahl, Alter, Zustand usw. den Bauwert beeinflussen, Bezirkseigenschaften dagegen den Bodenwert.

In dieser Arbeit wird daher sowohl der Haus- als auch der Bodenwert mit Hilfe der hedonischen Methode berechnet. Es wird insbesondere ermittelt, welche Eigenschaften eines Bezirks den Bodenwert beeinflussen, und wie hoch die impliziten Preise für diese Eigenschaften sind. Der Bodenwert ergibt sich analog zum Bauwert aus den Mengen an Charakteristika, die der Bezirk aufweist, und deren impliziten Preisen. Die hedonische Methode soll also zwei der Besonderheiten des Gutes Wohnen handhabbar machen: die Heterogenität der gehandelten Häuser und die Heterogenität der Bezirke, die wegen der Immobilität der Häuser einen wichtigen Einfluss auf den Preis hat.

Kapitel 3

Die Häuser

In dieser Arbeit wird der Kaufpreis eines bebauten Grundstücks in seine beiden Bestandteile zerlegt: in den Hausanteil, der durch die Eigenschaften des Hauses bestimmt wird, und in den Bodenanteil, der durch die Eigenschaften des Bezirks erklärt werden soll. Daher wird in diesem Kapitel zunächst der Hausdatensatz und dessen Modifikationen beschrieben, bevor im nächsten Kapitel zu den Bezirken übergegangen wird.

3.1 Der ursprüngliche Datensatz

Der Datensatz bestand ursprünglich aus 25.105 Beobachtungen und umfasste sämtliche Verkäufe von Einzelhäusern, Reihenhäusern und Doppelhaushälften im Zeitraum vom 01.01.1980 bis 31.07.2001. Hausverkäufe in Berlin-Ost wurden erst seit der Wiedervereinigung am 03.10.1990 erfasst.

Diese Daten werden vom Gutachterausschuss in Berlin gesammelt und zur Verfügung gestellt. Nach §193 Abs. 3 BauGB führt der Gutachterausschuss eine Kaufpreissammlung und wertet sie aus (vgl. Kapitel 2). Die Datenbasis deckt also sämtliche Transaktionen des Marktes für Einfamilienhäuser in Berlin ab.

Jede Beobachtung wird anhand verschiedener Variablen ausführlich beschrieben: Einige Merkmale sind quantitativ, d. h. ihre Merkmalsausprägungen sind Zahlen, wie z. B. **Grundstücksfläche**, **Geschossfläche** und **Brutto-rauminhalt**¹. Andere Variablen beschreiben eine Kategorie, der das betrachtete Haus angehört. Sie teilen die Daten anhand qualitativer Merkmale in zwei oder mehr Gruppen. Ein Haus kann z. B. anhand seines Zustandes (gut,

¹Variablen werden im folgenden durch **Schreibmaschinenschrift** gekennzeichnet.

normal, schlecht) oder seiner Blocklage (Frontgrundstück, Eckgrundstück, Hammergrundstück usw.) kategorisiert werden. Wiederum andere Variablen sind Indikator- oder Dummy-Variablen. Dummy-Variablen geben an, ob etwas wahr ist. Sie kommen nur in der Ausprägung 0 (Eigenschaft trifft nicht zu) und 1 (Eigenschaft trifft zu) vor. Diese Variablen geben direkt an, ob ein Charakteristikum vorhanden ist oder nicht. So gibt z. B. die Variable **Wasserlage** an, ob ein Grundstück am Wasser liegt oder nicht. Indikatorvariablen sind ein Spezialfall von kategoriellen Variablen. Die kategoriellen Variablen wurden nachträglich in Dummies umgewandelt, um für ihre Ausprägung in der Regression kontrollieren zu können. Beispielsweise wurde die Variable **Zustandsnote** mit den Ausprägungen “gut”, “normal”, und “schlecht” in die Dummies **note1** (gut), **note2** (normal) und **note3** (schlecht) aufgeteilt.

Die erfassten Variablen dienen verschiedenen Zwecken: Die erste Gruppe dient dazu, die Beobachtung zu identifizieren. Das geschieht anhand der Variablen **Nummer**, die jeder Beobachtung eine Nummer im Datensatz zuweist, **Straße**, **Grundstücksnummer**, **Grundstücksnummernzusatz**, **Ortsteil** und **Bezirk**, die gemeinsam den Standort des Objekts eindeutig identifizieren, sowie **Vertragsdatum**. Die letzte Variable ist wichtig, falls Häuser mehrfach verkauft wurden.

Die zweite Gruppe von Variablen beschreibt das Objekt umfassend. Hierzu zählen **Baujahr**, **Grundstücksfläche**, **Geschossfläche**, **Bruttorauminhalt**, **Haustyp** usw.

Die Gegend wird durch die dritte Gruppe von Variablen kurz charakterisiert, beispielsweise durch **Typische Nutzungsart**, **Typische Geschossflächenzahl** und **Wohnlage**. Da diese Variablen jedoch zum einen häufig nicht erfasst wurden, also viele missings aufweisen, und zum anderen bei weitem nicht ausreichen, um einen Bezirk ausführlich zu beschreiben, müssen zur Charakterisierung des Bezirks weitere Daten erfasst werden (s. Abschnitt 4.2).

Zur vierten Gruppe gehören sonstige preisbeeinflussende Faktoren wie **Ungewöhnliche Verhältnisse des Geschäftsverkehrs**, **Private Bindung**, **Ungewöhnliche Verhältnisse tatsächlicher Art** usw.

Obwohl das Hauptaugenmerk dieser Arbeit auf der Ermittlung von bodenpreisbeeinflussenden Bezirkseigenschaften liegt, können die Hauseigenschaften nicht vernachlässigt werden. So ist unmittelbar einsichtig, dass ein Haus auf einem großen Grundstück mehr kostet als auf einem kleinen und, wenn es in schlechtem Zustand ist, weniger als ein Haus in gutem Zustand. Für

diese Eigenschaften muss also kontrolliert werden.

Haurin und Brasington (1996) benutzen dafür z. B. das Alter des Hauses, die Quadratfußanzahl von Haus und Garage, die Anzahl der Badezimmer, Vorhandensein von Pool oder Klimaanlage usw. Des Rosiers et al. (2002) benutzen u. a. den Haustyp, das Alter, den Zustand und die Größe von Haus und Grundstück.

In dieser Arbeit wird für das Alter, die Grundstücks- und die Geschossfläche, Dachausbau, besondere Verhältnisse, Wasserlage, Blocklage, Dachform, Erwerbergruppe, Zustand, Stellung, Geschossanzahl, Veräußerergruppe und Verfügbarkeit kontrolliert. Im Anhang A Tabelle A.1 findet sich eine Liste der wichtigsten Hausvariablen und deren Erklärung.

3.2 Modifikationen des Datensatzes

Da in dieser Arbeit nur der Zeitraum 1991 bis 2000 betrachtet wird, wurden alle Beobachtungen, die nicht in diesen Zeitraum fallen, entfernt. Danach bestand der Datensatz aus 12.748 Beobachtungen.

Bei der Überprüfung des Datensatzes stellte sich heraus, dass zahlreiche Beobachtungen doppelt eingelesen wurden. Vier Beobachtungspaare waren völlig identisch. Die vier doppelt eingelesenen Beobachtungen wurden gelöscht.

1.470 Beobachtungen waren mit Ausnahme jeweils eines Merkmals paarweise identisch, und zwar dergestalt, dass bei einer ersten Erfassung ein Merkmal nicht erfasst wurde, dies bei einer zweiten Erfassung aber nachgeholt wurde, ohne jedoch die erste Erfassung zu löschen. Dies führte dazu, dass manche Beobachtungen doppelt auftraten und bis auf eine Merkmalsausprägung völlig identisch waren.

So gab es 733 Paare, die bis auf die Variable `Ortsteil` identisch waren. Die Beobachtungen, bei denen das Merkmal missing war, wurden gelöscht. Allerdings werden in dieser Arbeit nur Werte auf Bezirksebene betrachtet, so dass diese Variable für das weitere Vorgehen unwichtig ist.

Ähnliche Paare gab es für `Zulässige Geschossflächenzahl` (Anzahl der Paare: 1) und `Dachform` (1). Auch hier wurden die Beobachtungen mit der Ausprägung “missing” gelöscht.

Bei einem ansonsten identischen Beobachtungspaar wurde die Dachform unterschiedlich erfasst. Da nicht mehr nachvollziehbar war, welche die richtige ist, wurden beide Beobachtungen gelöscht.

Des weiteren wurden jeweils zwei Beobachtungen mit den Merkmalen Bogendach bzw. Sheddach und ein bauaufsichtlich gesperrtes Haus entfernt, da eine bzw. zwei Beobachtungen zu wenig sind, um für dieses Merkmal kontrollieren zu können.

Problematisch ist die Variable **Umfang**. Sie gibt den Erwerbsumfang bei ideellem Anteilserwerb an. Wenn dieses Feld belegt ist, beziehen sich alle Angaben mit Ausnahme des Kaufpreises jeweils auf das Gesamtgrundstück (vgl. Sofyan et al. (2002)) Dies war jedoch nicht immer der Fall. Da diese Beobachtungen unzuverlässig sind, wurden alle 497 Beobachtungen mit einem Erwerbsumfang von weniger als 100 % gelöscht.

Zwei weitere Beobachtungen wurden entfernt, da bei ihnen wahrscheinlich der Preis falsch erfasst wurde. Bei der ersten handelt es sich um ein DM 10 Millionen teures Reihenhauses im Bezirk Steglitz, Ortsteil Lichterfelde, in einfacher Lage, das von einer juristischen Person an eine andere juristische Person verkauft wurde. Hier ist es wahrscheinlicher, dass sich der Preis nicht auf ein einzelnes Reihenhauses bezieht, sondern auf ein Grundstück, auf dem mehrere Reihenhäuser stehen. Alle anderen Reihenhäuser in der betreffenden Straße kosteten deutlich unter DM 1 Million. Der durchschnittliche Preis für ein bebautes Grundstück betrug in Lichterfelde im Betrachtungszeitraum, ohne das betreffende Haus, DM 747.498. Bei der zweiten Beobachtung handelt es sich um eine mehr als DM 9 Millionen teure Doppelhaushälfte im Bezirk Treptow, Ortsteil Altglienicke, in mittlerer Lage. Im selben sowie im darauf folgenden Jahr wurden zwei Doppelhaushälften mit derselben Adresse, d. h. in derselben Straße mit derselben Grundstücksnummer, für DM 474.000 bzw. DM 485.000 verkauft, was dem Preisniveau der betrachteten Straße entspricht. Insgesamt hatte Treptow das niedrigste Preisniveau aller Bezirke. Der Preisdurchschnitt in Altglienicke war, ebenfalls ohne das betreffende Haus, DM 329.015. Eine Erklärungsmöglichkeit wäre, dass zunächst beide Doppelhaushälften gemeinsam gekauft wurden, der Preis sich also auf zwei Häuser bezieht, und zusätzlich bei der Erfassung der Kaufpreis mit einer Null zuviel angegeben wurde, der demnach für beide Doppelhaushälften zusammen gut DM 900.000 betrug. Kurz darauf wurden die beiden Doppelhaushälften einzeln mit Gewinn verkauft, für die oben angegebenen Preise. Da jedoch nicht nachvollziehbar ist, wie es sich tatsächlich verhalten hat und wie der Preis des Doppelhauses auf die beiden Doppelhaushälften aufgeteilt werden

soll (schließlich wurden sie auch zu verschiedenen Preisen wieder verkauft), wird der erste Verkauf des Doppelhauses aus dem Datensatz gestrichen. Übrig bleibt ein Datensatz mit 11.503 Beobachtungen.

Bei einigen Beobachtungen wurden kategorielle Variablen falsch in Dummies umgewandelt: So wies bei einer Beobachtung die **Dachform** die Ausprägung "Walmdach/Krüppelwalmdach" aus, alle zugehörigen Dachform-Dummies waren jedoch 0. Der entsprechende Dummy wurde auf 1 gesetzt. Bei einer Beobachtung wies die **Veräußerergruppe** die Ausprägung "privat" aus, die Dummies für die Ausprägungen "sonstiges Bundesland" und "Landeseigene Wohnungsunternehmen" waren jedoch missing gesetzt statt 0. Bei der Variable **Verfügbarkeit** mit der Ausprägung "nicht bezugsfertig" traten zwei Arten von Fehlern bei der Umwandlung in Dummy-Variablen auf: Bei 194 Beobachtungen waren alle Dummies 0, und bei 61 Variablen war der Dummy für "bezugsfrei" 1 gesetzt, der für "nicht bezugsfertig" jedoch auf 0. Alle fehlerhaften Dummies wurden korrigiert.

Bei den Variablen **Anzahl der Vollgeschosse** und **Typische Geschossflächenzahl** handelt es sich um diskrete Variablen. Daher wurden sie in Dummies umgewandelt. Da nur fünf Häuser vier Vollgeschosse hatten, wurden die Gruppen für Häuser mit drei Vollgeschossen und mit vier Vollgeschossen zu einer Gruppe zusammengefasst. Ebenso wurden bei der Variablen **Erwerbergruppe** folgende Ausprägungen zu einer Gruppe "sonstige" zusammengefasst: Bundesrepublik Deutschland (Anzahl Beobachtungen: 1), Land Berlin (2), sonstige öffentliche Hand (2), (gemeinnützige) Wohnungsunternehmen (1), Verfahrensträger (1), Versicherungsunternehmen (1), Kirche/Religionsgemeinschaft (1) und Diplomatische Vertretung (2). Bei der **Veräußerergruppe** wurden Immobilienfonds (2), Versicherungsunternehmen (3), Diplomatische Vertretung (2) und sonstiges Bundesland (1) zu einer Gruppe zusammengefasst. Die Ausprägungen "Dienstleistungsgewerbe" (5), "Produzierendes Gewerbe" (3) und "Mischgebietstypischer Charakter" (1) wurden bei der Variable **Typische Nutzungsart** zu einer Kategorie "sonstige Nutzungsart" zusammengefasst.

Kapitel 4

Die Bezirke

Da in dieser Arbeit untersucht werden soll, wie die Unterschiedlichkeit der einzelnen Bezirke den Preis von Immobilien beeinflusst, werden im ersten Abschnitt dieses Kapitels die Berliner Bezirke kurz beschrieben. Im zweiten Abschnitt wird überlegt, welche Bezirkseigenschaften den Kaufpreis beeinflussen könnten. Dafür wird die verfügbare Literatur zu Hilfe genommen. Allerdings kann nicht alles, was das Bild eines Bezirks beeinflusst, auch statistisch erfasst werden. Es wird daher auf jene Merkmale eingegangen, für die statistisches Datenmaterial vorliegt, und Hypothesen aufgestellt, wie diese den Kaufpreis beeinflussen könnten. Am Ende des Kapitels werden diese Hypothesen kurz zusammengefasst.

4.1 Beschreibung der Bezirke

Im betrachteten Zeitraum bestand Berlin aus 23 Bezirken. Mit der Bezirksfusion am 01.01.2001, also erst nach dem Betrachtungszeitraum dieser Arbeit, wurden diese zu zwölf Bezirken zusammengelegt. Daher werden in dieser Arbeit die 23 alten Bezirke analysiert. Von diesen liegen zwölf im Westteil der Stadt, nämlich Tiergarten, Wedding, Kreuzberg, Charlottenburg, Spandau, Wilmersdorf, Zehlendorf, Schöneberg, Steglitz, Tempelhof, Neukölln sowie Reinickendorf. Mit Mitte, Prenzlauer Berg, Friedrichshain, Treptow, Köpenick, Lichtenberg, Weißensee, Pankow, Marzahn, Hohenschönhausen und Hellersdorf gehören elf Bezirke zum Ostteil der Stadt. Im folgenden soll jeder Bezirk kurz charakterisiert werden. Dies erfolgt in Anlehnung an Plötz (2002) für die 23 alten Bezirke. Die Reihenfolge der Bezirke stimmt mit der vom Statistischen Landesamt Berlin für ihre Daten benutzten überein und wird in der gesamten Arbeit eingehalten.

- Mitte bildet das Handels-, Verwaltungs- und Regierungszentrum und

bietet zahlreiche kulturelle Attraktionen, wie z. B. zwei Opern, Theater, Bibliotheken und die Museumsinsel.

- In Tiergarten liegt die größte innerstädtische Parkanlage Berlins, die die Attraktivität der angrenzenden Wohnlagen erhöht. Sozial schwache Kieze gehören ebenso dazu wie das Diplomatenviertel, der Reichstag, das Bundeskanzleramt und Schloss Bellevue.
- Wedding zählt zu den am dichtesten besiedelten Stadtteilen und hat erhebliche Bedeutung als Gewerbestandort. Es überwiegen Altbauten in einfachen Wohnlagen.
- Prenzlauer Berg besteht überwiegend aus Altbau-Miethäusern. Dieser Stadtteil ist wegen seiner Kneipen- und Kulturszene sehr beliebt.
- Friedrichshain ist beliebt bei jungen Leuten und Studenten wegen seiner Szene und der multikulturellen Vielfalt. Die Wohnlagen sind einfach und mittel.
- Auch in Kreuzberg herrscht multikulturelle Vielfalt. Schwache Sozialstrukturdaten, niedrige Einkommen und hohe Arbeitslosigkeit machen Kreuzberg zu einem der Problemstadtteile Berlins.
- Der Kernbereich von Charlottenburg bildet den wichtigsten Teil der "City West" mit zahlreichen Shopping-Möglichkeiten. Hier finden sich einige der teuersten Wohnlagen. Es gibt ein großes Angebot an Kinos, Restaurants, Kneipen, Theater, Museen und eine Oper.
- Spandau ist der größte Industriestandort der Stadt. Den Kern bildet die mittelalterliche Altstadt Spandau. Die Bebauung besteht sowohl aus Einfamilien- als auch aus Hochhäusern. Im Norden und Süden gibt es große Wald- und Wasserflächen.
- Wilmersdorf wird zur Hälfte vom Grunewald und von Wasserflächen eingenommen. Es gibt Einfamilienhaus- und Villengebiete, aber auch dichter besiedelte Gebiete, die zur Innenstadt gehören.
- Zehlendorf gilt als nobelste Wohngegend und ist von Villen sowie Ein- und Zweifamilienhäusern geprägt. Über 50 % der Fläche Zehlendorfs ist von Wald und Wasser bedeckt und bietet damit einen hohen Naherholungswert.
- In Schöneberg finden sich zahlreiche Altbauten. Es gibt viele Gewerbebetriebe, ein Teil des Bezirks gehört zur "City-West". Die Wohnlagen sind einfach bis sehr gut.

- In Steglitz finden sich Wohngebiete mit gutbürgerlichem Ambiente. Im Süden gibt es Villen- sowie Ein- und Zweifamilienhausgebiete.
- Im Norden Tempelhofs gibt es größere Altbaubestände, während sich im Süden ausgedehnte Einfamilienhausgebiete befinden. Nach Spandau ist Tempelhof der zweitgrößte Industriestandort Berlins.
- Neukölln gehört ebenfalls zu den wichtigsten Gewerbestandorten. Fast ein Viertel des Bezirks besteht aus Grünflächen.
- Treptow: die Wohnlagen sind von einfacher bis mittlerer Qualität. In Altglienicke finden sich Plattenbauten, am Stadtrand Einfamilienhausgebiete.
- Köpenick ist der grünste und der am dünnsten besiedelte Bezirk (vgl. Abschnitt 6.2). Es finden sich alte Villenkolonien. Die Wohnlagen sind einfach bis gut.
- Lichtenberg wird geprägt von Plattenbauten, aber es findet sich hier auch die alte Villenkolonie Karlshorst. Es gibt zahlreiche Einkaufsmöglichkeiten sowie Schulen und Kindergärten.
- Weißensee: Zwei Drittel der Wohnungen entstanden vor 1946, ein Drittel nach 1990. Es gibt zahlreiche kleine Gewerbebetriebe und ausgedehnte Gebiete mit Ein- und Zweifamilienhäusern. Die Wohnlagen sind einfach bis gut.
- In Pankow gibt es Mietshäuser, überwiegend Altbauten, kleinere Villenviertel und ausgedehnte Einfamilienhausgebiete in mittleren und guten Wohnlagen.
- Reinickendorf ist großzügig bebaut und bietet mit seinen Wald- und Wasserflächen eine hohe Wohnqualität. Die Wohnlagen sind mittel bis gut, teilweise auch sehr gut.
- Marzahn: Auch in diesem Bezirk überwiegen Plattenbauten, einzelne Ortsteile haben Gartenstadtcharakter. Die Bevölkerung umfasst alle sozialen Schichten.
- Hohenschönhausen wird ebenfalls von Plattenbauten geprägt, von denen viele leerstehen. Zu dem Bezirk gehören aber auch Altbauviertel und ehemalige Dörfer wie Falkenberg, Malchow und Wartenberg.

- Hellersdorf: Die Stadtteile Biesdorf (Marzahn), Mahlsdorf und Kaulsdorf bilden das größte zusammenhängende Einfamilienhausgebiet Berlins.

In der Abbildung 4.1 sind sowohl die 23 alten Bezirke als auch die zwölf neuen Bezirke Berlins eingezeichnet.



Abbildung 4.1: Neue und alte Bezirke Berlins. Quelle: Statistisches Landesamt Berlin.

4.2 Kaufpreisbeeinflussende Bezirkseigenschaften

Bei der Entscheidung, wo das zu kaufende Objekt liegen soll, spielen verschiedene Aspekte eine Rolle. Einer davon ist sicher der Traum von einem "Haus im Grünen". Ein Bezirk mit vielen Parkflächen, Wäldern und Seen müsste also einem stark bebauten oder von Gewerbe geprägten Bezirk vorgezogen

werden. Wie grün ein Bezirk ist, kann z. B. anhand der Variable Stadtfläche, die aus den Teilflächen bebaute Fläche, Betriebsfläche, Erholungsfläche, Verkehrsfläche, Landwirtschaftsfläche, Waldfläche, Wasserfläche und andere Flächen besteht, sowie anhand der Anzahl der Straßenbäume auf einen Kilometer Straßenlänge gemessen werden. Der Einfluss von Bäumen und Wald auf Hauspreise wurde z. B. von Theriault et al. (n.d.), Des Rosiers et al. (2002) und Luttik (2000) untersucht. Des Rosiers et al. (2002) finden einen positiven Effekt auf Hauspreise, wenn das Grundstück stärker mit Bäumen bepflanzt ist als die Nachbarschaft, solange der Unterschied nicht zu groß ist. Der Effekt wird umso stärker, je mehr ältere Personen über 65 Jahren dort leben. Ebenso haben Rasen und Blumen einen positiven Effekt, je nach Haustyp. Auch Luttik (2000) findet positive Effekte. Dagegen finden Theriault et al. (n.d.) einen vom sozio-ökonomischen Status der Nachbarschaft abhängigen Effekt, der in beide Richtungen gehen kann. In dieser Arbeit werden die Teilflächen der Stadtflächen untereinander und zu der Bevölkerung in Beziehung gesetzt. Leider wurden die Wohnfläche und die Gewerbefläche in den Bezirken Prenzlauer Berg, Treptow, Köpenick, Lichtenberg und Marzahn ab 1995 nicht mehr erfasst, so dass deren Einfluss nicht untersucht werden konnte.

Weiterhin wird in der Literatur häufig die Erreichbarkeit des "Central Business Districts (CBD)" gemessen. Unterschiede im Bodenpreis werden auf Unterschiede in den Transportkosten zum CBD zurückgeführt (s. z. B. Haurin und Brasington (1996)). Wie Case und Mayer (1996) betonen, spielt bei der Erreichbarkeit der nächsten Großstadt nicht nur der Zugang zu Jobs eine Rolle, sondern auch der Zugang zu kulturellen Aktivitäten wie Theater, Restaurants, Sportereignissen usw. Dies soll in dieser Arbeit so keine Rolle spielen, da sowohl die Arbeitgeber als auch kulturelle Plätze über die gesamte Stadt verteilt sind. Allerdings wird die Qualität der Verkehrsanbindung der einzelnen Bezirke untersucht. Diese soll anhand der Variablen S- und U-Bahnstationen im Bezirk gemessen werden. Diese Daten wurden aus Angaben der S-Bahn Berlin GmbH und der Berliner Verkehrsbetriebe (BVG), die für die U-Bahnen zuständig sind, zusammengestellt. Hier ist allerdings zu beachten, dass an einer S- oder U-Bahnstation mehrere S- bzw. U-Bahnlinien halten können. Als Spitzenreiter ist der S-Bahnhof Westkreuz im Bezirk Charlottenburg zu nennen, an dem neun S-Bahnlinien halten. Daher wäre es sicher aussagekräftiger, wenn die Bahnhöfe mit der Anzahl der dort haltenden Linien gewichtet werden könnten. Leider konnten weder die S-Bahn Berlin GmbH noch die BVG entsprechende Zahlen nennen. Beide stellten nur jeweils eine Liste mit den aktuellen S- bzw. U-Bahnhöfen bereit mit Hinweisen darauf, wann welche Streckenabschnitte eröffnet wurden. Je-

der Bahnhof wurde einem Bezirk zugeordnet und vom heutigen Stand die Anzahl der Bahnhöfe bis zum Jahr 1991 rückgerechnet. Dies gestaltete sich für die Anzahl der Bahnhöfe relativ problemlos, war jedoch weitgehend unlösbar für die Anzahl der haltenden Linien an den einzelnen Bahnhöfen, weil sich der Linienverlauf innerhalb des letzten Jahrzehnts mehrfach geändert hat und daher nicht mehr nachvollziehbar war. Es wird daher nur der Einfluss der Anzahl der S- und U-Bahnstationen in einem Bezirk untersucht. Die Hypothese lautet, dass der Preis für ein bebautes Grundstück steigt, je besser die bereitgestellte Verkehrsinfrastruktur ist, d. h. je mehr S- und U-Bahnstationen es in einem Bezirk gibt.

Nicht nur die Verkehrsanbindung, sondern auch Einkaufsmöglichkeiten sind Teil der Infrastruktur eines Bezirks. Den Einfluss der Größe von und der Nähe zu Einkaufszentren auf Hauspreise haben z. B. Des Rosiers et al. (1996) untersucht. Einkaufszentren können positive Effekte durch bequeme Erreichbarkeit von Einkaufs- und Unterhaltungsmöglichkeiten haben, aber auch negative externe Effekte durch Lärm, Verkehrsstau und Umweltverschmutzung. Die Autoren stellen einen positiven Einfluss für die Größe fest, aber eine Nicht-Monotonität bei der Preis-Entfernung-Beziehung. Für verschiedene Arten von Einkaufszentren werden verschiedene optimale Entfernungen berechnet. Leider waren Angaben zu Ladenflächen in den einzelnen Bezirken nicht erhältlich.

In zahlreichen Studien wird der Einfluss der Qualität von Schulen auf Hauspreise untersucht (z. B. Downes und Zabel (1997), die eine ausführliche Übersicht über die entsprechende Literatur geben, Haurin und Brasington (1996), Rosenthal (2000) und Walden (1990)). Die Qualität der Schule wird dabei z. B. anhand der Größen Schüler-Lehrer-Verhältnis, durchschnittliche Erfahrung der Lehrer, öffentliche Ausgaben pro Schüler und Ergebnisse landesweiter Tests gemessen. Da diese Angaben hier nicht vorliegen, soll statt dessen als ein grobes Maß für die Klassengröße zum einen der Einfluss der Anzahl der Kinder unter sechs Jahren pro Grundschulklasse und zum anderen der Einfluss der Kinder von sechs bis unter 15 Jahren pro Grundschulklasse untersucht werden. Im ersten Fall wird unterstellt, dass die Anpassung der Kapazitäten zeitverzögert erfolgt, dass also die zukünftig einzuschulenden Kinder in der von der jetzigen Schülergeneration beeinflussten Anzahl von Klassen sitzen und für Käufer mit Kindern unter sechs Jahren die heutige Schullandschaft entscheidend ist. Es wird die zukünftige Klassengröße betrachtet. Im zweiten Fall wird dagegen die heutige Klassengröße betrachtet. Zu beachten ist allerdings, dass der Übergang zur Oberschule in Berlin mit dem zwölften oder dreizehnten Lebensjahr erfolgt, so dass die heutige Klassengröße

überschätzt wird. Die Analyse beschränkt sich nur auf Grundschulen, weil die Grundschule besucht werden muss, von Ausnahmen abgesehen, “in deren Einschulungsbereich das Kind seinen gewöhnlichen Aufenthalt hat (zuständige Grundschule)” (§8 (3) SchulG). Für weiterführende Schulen gilt diese Vorschrift nicht, so dass auch Oberschulen außerhalb des Bezirks besucht werden können. Mit dieser Variablen wird ein Teil der sozialen Infrastruktur gemessen. Analog zur Verkehrsinfrastruktur lautet die Hypothese, dass der Preis bei besserer sozialer Infrastruktur, d. h. mit abnehmender Klassengröße, steigt.

Obwohl die Anzahl der Spielplätze und die Spielfläche ebenfalls zur sozialen Infrastruktur eines Bezirks gehören, soll hier eine andere Hypothese untersucht werden. Es könnte sein, dass Spielplätze vor allem dort eingerichtet werden, wo Miethäuser das Stadtbild beherrschen und es wenig Grünflächen oder gar einen eigenen Garten gibt, in dem Kinder spielen können. Kinderspielplätze könnte es daher vor allem in den billigen Bezirken geben. Die Hypothese lautet daher, dass der Preis umso geringer ist, je größer die Spielfläche in einem Bezirk ist.

Ferner spielen sozio-ökonomische Aspekte eine Rolle. Dazu gehören u. a. das Einkommen, die Arbeitslosigkeit und die Anzahl der von den Erwerbspersonen zu unterhaltenden Nichterwerbspersonen. Käufer, die bereit und in der Lage sind, viel Geld für ein Objekt auszugeben, schätzen wahrscheinlich Bezirke mit wohlhabender Bevölkerung und geringer Arbeitslosigkeit. Walden (1990) begründet den von ihm gefundenen positiven Effekt eines steigenden Medianeinkommens mit positiven Externalitäten, die Hauskäufer mit wohlhabenden Gegenden assoziieren. Auch das Bildungsniveau eines Bezirks könnte einen positiven Einfluss haben. Downes und Zabel (1997) untersuchen beispielsweise den Einfluss des Anteils der Bevölkerung über 25 Jahren mit High-School-Abschluss, ohne jedoch einen signifikanten Einfluss festzustellen. Leider gibt es auch dazu keine Daten. Unterstellt man aber, dass jemand mit einem höheren Bildungsniveau im Schnitt mehr verdient als jemand mit einem geringeren, so misst das Nettoeinkommen auch das Bildungsniveau mit.

Das Bild eines Bezirks wird auch durch seine ethnische Zusammensetzung bestimmt. Zahlreiche anglo-amerikanische Studien beschäftigen sich mit der Segregation aufgrund der Rasse und deren Gründen und Auswirkungen. Als Beispiele seien nur die Studien von Bajari und Kahn (2002) und Bayer und McMillan (2001) genannt. Bajari und Kahn (2002) kommen zu dem Ergebnis, dass Schwarze und Weiße deswegen in unterschiedlichen Teilen der Stadt wohnen - nämlich die Schwarzen in der Innenstadt und die Weißen in den

Vororten -, weil sie unterschiedliche Nachfragen haben, die sich angleichen würden, wenn sie dasselbe Einkommen und denselben Bildungsstand hätten. Dagegen lehnen sie die Hypothese ab, dass Weiße deswegen in den Vororten wohnen, weil sie unter anderen Weißen leben wollen. Ein hoher Ausländeranteil könnte daher auf billige Bezirke hindeuten. Bayer und McMillan (2001) erwähnen die negativen Auswirkungen für ein Individuum bei extremer Entmischung der Rassen, wie sie in Ghettos zu finden ist, ausgelöst durch Armut und negative externe Effekte. Cutler und Glaeser (1995) weisen darauf hin, dass Segregation auch positive Effekte haben könnte, da sie u. a. zu homogenen Gruppen mit identischen Nachfragen nach öffentlichen Gütern führt.

Außerdem wird der Einfluss der Altersstruktur eines Bezirks anhand des Anteils an Kindern bis unter 15 Jahren an der Bevölkerung eines Bezirks untersucht. Geht man davon aus, dass diejenigen, die genügend Geld verdienen, so dass sie zwischen allen möglichen Bezirken auswählen können und nicht aufgrund ihres Einkommens auf bestimmte Bezirke beschränkt sind, eher wenige Kinder haben, so könnte der Anteil an Kindern einen negativen Einfluss haben, da sich eben jene Leute von Kindern beeinträchtigt fühlen könnten. Stellen dagegen Familien mit Kindern die Hauptnachfrager, dann hätte der Anteil der Kinder einen positiven Effekt, da diese Familien eventuell Spielkameraden für ihre Kinder und gleichgesinnte Nachbarn suchen. Walden (1990) stellt beispielsweise einen Preisanstieg fest, wenn der Anteil der Bevölkerung über 65 oder unter 18 Jahren steigt, ohne eine Erklärungsmöglichkeit dafür zu liefern. Downes und Zabel (1997) finden einen positiven Einfluss, wenn der Median des Alters der Bevölkerung steigt. Die hier zu überprüfende Hypothese lautet also in Anlehnung an die Ergebnisse von Walden (1990), dass der Anteil an Kindern in einem Bezirk die Hauspreise positiv beeinflusst.

Eine hohe Kriminalitätsrate wird die Kaufentscheidung wahrscheinlich negativ beeinflussen. Gibbons (n.d.) untersucht den Einfluss von Eigentumsdelikten auf Hauspreise in London unter Beachtung möglicher Endogenität der Kriminalitätsrate. Er findet keinen messbaren Einfluss von Einbrüchen auf Immobilienpreise, dagegen einen negativen Einfluss von Sachbeschädigungsdelikten wie Vandalismus und Graffiti. Diese könnten als Signale für Instabilität und Verschlechterung der Nachbarschaft gelten und die Furcht vor Kriminalität erhöhen, obwohl er nur eine schwache Korrelation dieser Delikte mit schweren Delikten feststellt. Deller und Ottem (2001) untersuchen den Einfluss verschiedener Straftaten auf Hauspreise eingehend auf drei Spezifikationsstufen. Auf der ersten wird ein allgemeiner Kriminalitätsindex benutzt und ein negativer Einfluss auf Hauspreise festgestellt. Auf der zweiten Stufe werden die Straftaten in Eigentumsdelikte und Gewaltverbrechen

aufgeteilt. Eigentumsdelikte haben einen negativen Effekt, Gewaltverbrechen dagegen einen insignifikanten oder positiven, je nachdem, wie die Schätzung vorgenommen wird. Auf der dritten Stufe werden die Eigentumsdelikte und Gewaltverbrechen in jeweils vier Kategorien eingeteilt. Die Koeffizienten für Vergewaltigung, Einbruch und leichter Diebstahl sind negativ, die für schwere Körperverletzung und Autodiebstahl positiv sowie die für Mord, Raub und Brandstiftung insignifikant. Sie weisen darauf hin, dass eine hohe Kriminalitätsrate einen Bezirk unattraktiv machen kann, dass aber im Gegenzug wohlhabendere Gegenden bestimmte Arten von Kriminalität anziehen. Leider wird die Polizeiliche Kriminalstatistik, in der die der Polizei bekannt gewordenen Straftaten nach Abschluss der polizeilichen Ermittlungen registriert werden, erst seit 2001 auf Bezirksebene erfasst. Vorher wurde Berlin insgesamt als eine Region gesehen. Diese Daten können daher für die vorliegende Arbeit nicht genutzt werden.

In der Literatur werden die Aspekte nicht einzeln untersucht. Beispielsweise benutzt Walden (1990) in seiner Untersuchung über den Einfluss der Schulqualität auf den Wert eines bebauten Grundstücks den Steuersatz, den Median des Haushaltseinkommens, den Anteil der Bevölkerung über 65 und unter 18 Jahren sowie den Anteil der nicht-weißen Bevölkerung, um die Nachbarschaft oder Gegend zu beschreiben. Haurin und Brasington (1996) benutzen das Durchschnittseinkommen, den Anteil weißer Haushalte sowie die Kriminalitätsrate.

4.3 Zusammenfassung der Hypothesen

Zusammenfassend ergeben sich folgende Arbeitshypothesen: Der Preis eines bebauten Grundstücks ist umso höher,

1. je "grüner" ein Bezirk,
2. je besser die Verkehrsanbindung,
3. je geringer die heutige bzw. zukünftige Klassengröße von Grundschulklassen,
4. je weniger Kinderspielplätze es gibt,
5. je höher das Haushaltsnettoeinkommen,
6. je geringer die Arbeitslosigkeit,

7. je weniger Nichterwerbspersonen auf eine Erwerbsperson kommen,
8. je geringer der Ausländeranteil und
9. je höher der Anteil an Kindern ist.

Kapitel 5

Statistische Grundlagen

In dieser Arbeit wird untersucht, durch welche Größen der Preis eines Hauses beeinflusst wird. Der Zusammenhang zwischen einer zu erklärenden Größe, in dieser Arbeit der Preis eines bebauten Grundstücks, und ihren Einflussgrößen, hier die verschiedenen Haus- und Bezirkseigenschaften, wird empirisch mit Hilfe von Regressionsmodellen überprüft. Die Kenntnis statistischer Grundlagen ist von größter Wichtigkeit für das Verständnis des Analyseteils. Daher wird zunächst die hier verwendete hedonische Methode beschrieben. Dabei handelt es sich um einen Sonderfall des allgemeinen linearen Regressionsmodells. Dieses wird im zweiten Abschnitt des Kapitels kurz erklärt.

5.1 Hedonische Regression

Die hedonische Methode geht davon aus, dass ein Gut durch einen Vektor von K objektiv messbaren Charakteristika oder Attributen (x) beschrieben wird, d. h. $x = (x_1, x_2, \dots, x_K)$, wobei x_k die Menge des k -ten Charakteristikums misst, die in jedem Gut enthalten ist (vgl. Rosen (1974)).

Der Preis dieses Gutes zu einem bestimmten Zeitpunkt ist eine Funktion seiner Qualitätscharakteristika, d. h. $p(x) = p(x_1, x_2, \dots, x_K)$ (vgl. Rosen (1974)). Aus verschiedenen Versionen des Gutes mit jeweils unterschiedlichen Niveaus bestimmter Charakteristika kann ein impliziter oder hedonischer Preis für jedes Charakteristikum abgeleitet werden (vgl. Griliches und Ohta (1976)), indem der Preis eines bestimmten Modells eines Gutes gegen die verschiedenen Mengen an Charakteristika regressiert wird (vgl. Gordon (1990)). Hedonische Preise sind demnach definiert als die impliziten Preise von Attributen und werden dem ökonomischen Agenten durch die beobachteten Preise differenzierter Güter und die spezifischen Mengen an Charakte-

ristika, die mit diesen assoziiert sind, offenbart (vgl. Rosen (1974)).

Die erste empirische Untersuchung, die den Effekt von Qualität auf den Preis untersucht, wurde schon 1927 von Frederick Waugh durchgeführt, der den Einfluss von Qualität auf Gemüsepreise analysiert (vgl. Berndt (1991)). Court (1939) und Griliches (1961) haben die Methode für die Konstruktion von Preisindizes für Automobile benutzt. Ein wichtiger Punkt ist dabei, wie Preisindizes an Qualitätsänderungen angepasst werden sollen (vgl. Berndt (1991)). Preisindizes sollen den reinen Preisanstieg von Gütern konstanter Qualität ausweisen. Im Laufe der Zeit verbessert sich jedoch die Qualität der Produkte. Dies kann zum Ausweis eines Preisanstiegs führen, der jedoch ursächlich auf die verbesserte Qualität zurückzuführen ist und sich in einen Ausweis gesunkener Preise verwandelt, sobald man den Qualitätsverbesserungen Rechnung trägt. Moulton (2001) führt den Ursprung der offiziellen US Preisstatistik auf Griliches' Artikel zurück. Rosen (1974) hat mit seinem klassischen Artikel die theoretischen Grundlagen geschaffen. Er untersucht innerhalb eines mikroökonomischen Rahmens die Angebots- und Nachfrageseite und führt diese zu einem Marktgleichgewicht zusammen. Bis heute wurde die Methode in zahlreichen wirtschaftlichen Bereichen angewandt, besonders auf Automobile und Computer, aber auch auf Immobilien.

Dabei ist zwischen Arbeiten, die mit Hilfe der hedonischen Methode Preisindizes entwickeln, die Qualitätsänderungen im Zeitablauf und zwischen verschiedenen Regionen Rechnung tragen, und Arbeiten, die den Einfluss einzelner Qualitätscharakteristika auf den Preis untersuchen, zu unterscheiden. Moulton (2001) gibt einen geschichtlichen Überblick über die Preisstatistik der Vereinigten Staaten und die Rolle der hedonischen Methode. Im Bereich von Immobilien seien nur als Beispiel die Arbeiten von Moulton (1995) und Sieg et al. (n.d.) genannt. Diewert (2001) reduziert die Komplexität der dahinterstehenden Theorie durch vereinfachende Annahmen und diskutiert mögliche Formen der Funktion. Die Literatur für die zweite Gruppe von Untersuchungen wurde bereits an den entsprechenden Stellen dieser Arbeit genannt.

5.2 Das allgemeine lineare Regressionsmodell

5.2.1 Form der Funktion

Beim allgemeinen linearen Regressionsmodell wird davon ausgegangen, dass eine lineare Beziehung zwischen einer abhängigen Variablen, y , und K unab-

hängigen x -Variablen besteht:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_K x_{iK} + \epsilon_i \text{ mit } i = 1, 2, \dots, n \text{ und } k = 1, 2, \dots, K \quad (5.1)$$

wobei

- x_{i1} = 1 für alle i
- β_k - Regressionskoeffizienten, $k = 1, 2, \dots, K$ und
- β_1 - Regressionskonstante
- n - Umfang der Grundgesamtheit (Anzahl Beobachtungen)
- ϵ - Störgröße
- i - Index für Beobachtungen
- k - Index für Variablen

In Matrixnotation lauten die n Beobachtungsgleichungen

$$y = X\beta + \epsilon \quad (5.2)$$

mit

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} x_{11} & \dots & x_{1K} \\ \vdots & & \vdots \\ x_{nK} & \dots & x_{nK} \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_K \end{bmatrix}, \epsilon = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

(Vgl. Hübler (1989) und Judge et al. (1998).)

Verschiedene nichtlineare Beziehungen lassen sich durch semi- oder doppelt-logarithmische Transformation linearisieren. So wird z. B. aus $y = x_1^{\beta_1} \cdot x_2^{\beta_2} \cdot \dots \cdot \exp\{\epsilon_i\}$ durch doppeltlogarithmische Transformation

$$\ln y = \beta_1 \ln x_1 + \beta_2 \ln x_2 + \dots + \epsilon_i \quad (5.3)$$

und aus $y = a \exp\{\beta_1 x_1 + \beta_2 x_2 + \dots + \epsilon_i\}$ durch semilogarithmische Transformation

$$\ln y = \ln a + \beta_1 x_1 + \beta_2 x_2 + \dots + \epsilon_i \quad (5.4)$$

(vgl. Hübler (1989)). Gleichung 5.3 wird auch als log-log-Modell und Gleichung 5.4 als log-lineares Modell bezeichnet.

Dabei ist zu beachten, dass nur metrisch skalierte Variablen logarithmiert werden können, nicht jedoch Dummy-Variablen, die nur die Ausprägungen $[0, 1]$ annehmen können, da der Logarithmus von 0 nicht definiert ist.

5.2.2 Dummies und kategorielle Variablen

Dummy-Variablen können nur zwei Ausprägungen annehmen:

$$D = \begin{cases} 0 & \text{wenn Eigenschaft nicht zutrifft} \\ 1 & \text{wenn Eigenschaft zutrifft} \end{cases}$$

Die Dummy-Variable teilt also den Datensatz in zwei Gruppen.

Das allgemeine lineare Regressionsmodell kann dann geschrieben werden als

$$y_i = \beta_1 x_{i1} + \delta D_i + \beta_2 x_{i2} + \dots + \beta_K x_{iK} + \epsilon_i. \quad (5.5)$$

Durch den Term δD_i ist der y-Achsenabschnitt in den beiden Gruppen unterschiedlich (beachte, dass $x_{i1} = 1$ für alle i):

$$E[y_i] = \begin{cases} \beta_1 + \beta_2 x_{i2} + \dots + \beta_K x_{iK} & \text{wenn } D = 0 \\ (\beta_1 + \delta) + \beta_2 x_{i2} + \dots + \beta_K x_{iK} & \text{wenn } D = 1 \end{cases}$$

Für die zweite Gruppe von Daten ist der y-Achsenabschnitt $(\beta_1 + \delta)$ statt β_1 (vgl. Judge et al. (1998)). Gibt es nur zwei mögliche Zustände, so reicht eine Dummy-Variable.

Kategorielle Variablen teilen den Datensatz anhand qualitativer Merkmale in zwei oder mehr Gruppen. "Als eine generelle Regel gilt, daß, sofern eine qualitative Variable in eine von m Klassen fallen kann, zusätzlich zum Ordinatenabschnitt $m - 1$ Dummyvariablen notwendig sind" (Harvey; 1994). Der nicht modellierte Zustand wird als Referenzzustand benutzt. Zur Interpretation des Referenzzustandes s. Abschnitt 5.2.9.

Zusätzlich zu den Dummies für qualitative Variablen werden auch Dummies für jedes Jahr eingeführt. Damit kann für verschiedene Preisniveaus in den einzelnen Jahren kontrolliert werden, deren Ursachen nicht in einer Veränderung der Bezirksvariablen liegen, sondern alle Bezirke gemeinsam betreffen, wie z. B. das Zinsniveau.

Bei der Regressionsanalyse können Dummies wie alle anderen Variablen behandelt werden. Es ergeben sich keine Besonderheiten.

5.2.3 Annahmen

Die Annahmen des allgemeinen linearen Regressionsmodells lauten wie folgt:

Es soll keine Multikollinearität vorliegen, d. h. die Spaltenvektoren von X sind linear unabhängig, so dass sich kein Vektor x_k als Linearkombination der anderen darstellen lässt. Daraus folgt, dass

$$rg(X) = K.$$

X besitzt also den vollen Spaltenrang, und damit ist $X'X$ nichtsingulär und invertierbar. Diese Annahme ist wichtig, um den Kleinsten-Quadrate-Schätzer (KQ-Schätzer) für β nach Gleichung 5.6 berechnen zu können.

Der Erwartungswert der Störgröße soll Null sein, d. h.

$$E[\epsilon_i] = 0$$

und

$$E[\epsilon\epsilon'] = E \begin{bmatrix} \epsilon_1\epsilon_1 & \epsilon_1\epsilon_2 & \dots & \epsilon_1\epsilon_n \\ \vdots & & & \vdots \\ \epsilon_n\epsilon_1 & \epsilon_n\epsilon_2 & \dots & \epsilon_n\epsilon_n \end{bmatrix} = \begin{bmatrix} \sigma^2 & & & 0 \\ & \sigma^2 & & \\ & & \ddots & \\ 0 & & & \sigma^2 \end{bmatrix}_{(n \times n)} = \sigma^2 I_n.$$

Daraus folgt, dass die Kovarianzen null sind, also keine Autokorrelation vorliegt, und die Varianzen identisch sind, also Homoskedastie vorliegt.

5.2.4 Parameterschätzung nach der Methode der kleinsten Quadrate

Im Grundmodell nach Gleichung 5.1 sind y_i und x_{ik} beobachtbar, β_k und ϵ_i jedoch nicht. Es müssen daher Schätzer für $(\beta_1, \dots, \beta_K)$ und σ gefunden werden. Bei der Methode der kleinsten Quadrate (auch OLS für ordinary least squares genannt) sollen die Koeffizienten so geschätzt werden, dass die geschätzte Quadratsumme der Störgrößen minimiert wird:

$$\text{Min. } \sum_{i=1}^n \epsilon_i^2.$$

In Matrixnotation:

$$\text{Min. } \epsilon'\epsilon = S = (y - X\beta)'(y - X\beta).$$

Wird dieser Ausdruck nach β abgeleitet und die 1. Ableitung gleich Null gesetzt, so ergibt sich nach einigen Umformungen als KQ-Schätzer:

$$\hat{\beta} = (X'X)^{-1}X'y. \quad (5.6)$$

Der Schätzer wird in Abgrenzung zum wahren (unbekannten) Parameter β mit $\hat{\beta}$ bezeichnet.

Der Ansatz für die Schätzung der Störgrößenvarianz σ^2 lautet:

$$y - \hat{y} = \hat{\epsilon}.$$

Allerdings sind weniger die Residuen selbst von Interesse als die Residuenquadratsumme $\hat{\epsilon}'\hat{\epsilon} = (y - \hat{y})'(y - \hat{y})$. Weiterhin gilt $\hat{\epsilon}'\hat{\epsilon}/(n - K) = \hat{\sigma}^2$. Daraus folgt:

$$\hat{\sigma}^2 = \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{n - K}. \quad (5.7)$$

Eine ausführliche Herleitung würde hier zu weit führen. Es sei daher auf die weiterführende Literatur verwiesen, z. B. Hübler (1989) und Judge et al. (1998).

5.2.5 Eigenschaften des Kleinste-Quadrate-Schätzers

Der Kleinste-Quadrate-Schätzer $\hat{\beta}$ hat folgende Eigenschaften:

1. Der KQ-Schätzer $\hat{\beta}$ ist eine Funktion von y und damit eine Zufallsvariable.
2. Er ist linear in y .
3. Er ist erwartungstreu oder unverzerrt, d. h. $E[\hat{\beta}] = \beta$.
4. Die Varianz-Kovarianzmatrix lautet:

$$Cov[\hat{\beta}] = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = \sigma^2(X'X)^{-1}.$$

5. $\hat{\beta}$ ist effizient, d. h. BLUE (best linear unbiased estimator), wobei "best" bedeutet, dass er die kleinste Varianz unter allen linearen und unverzerrten Schätzern aufweist. Diese Aussage wird als Gauss-Markov-Theorem bezeichnet.

$\hat{\sigma}^2$ ist ebenfalls ein unverzerrter Schätzer für σ .

Für eine ausführliche Darstellung der Eigenschaften und einen Beweis des Gauss-Markov-Theorems vgl. Hübler (1989) und Judge et al. (1998).

5.2.6 Parameterschätzung nach der Maximum-Likelihood-Methode

Für die Maximum-Likelihood-Methode (ML-Methode) wird die Annahme über die Verteilung des Residuums ϵ geändert. Während sie vorher lautete, dass $\epsilon \sim (0, \sigma^2 I_n)$ (vgl. Abschnitt 5.2.3), lautet sie jetzt $\epsilon \sim N(0, \sigma^2 I_n)$. ϵ soll nun ein Vektor normalverteilter Zufallsvariablen sein. Daraus folgt, dass y ein multivariat normalverteilter Zufallsvektor ist: $y \sim N(X\beta, \sigma^2 I_n)$. Die Dichtefunktion für den Störgrößenvektor lautet

$$f(\epsilon) = \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta) \right).$$

Wenn die Störgrößen unabhängig identisch normalverteilt sind, gilt

$$f(\epsilon_i) = \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2} \epsilon_i^2 \right).$$

Daraus folgt für die logarithmierte Likelihoodfunktion

$$\ln L = \sum_{i=1}^n \ln f(\epsilon_i) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta).$$

Nun werden die 1. Ableitungen nach β und σ gebildet und gleich Null gesetzt.

Der Schätzer für β lautet wie der KQ-Schätzer

$$\hat{\beta}_{ML} = \hat{\beta}_{KQ} = (X'X)^{-1} X'y. \quad (5.8)$$

Dieser Schätzer ist BUE, d. h. er ist der beste unverzernte Schätzer, nicht nur auf die Klasse der linearen Schätzer beschränkt.

Der Schätzer für σ lautet

$$\hat{\sigma}_{ML}^2 = \frac{(y - X\hat{\beta}_{ML})'(y - X\hat{\beta}_{ML})}{n}, \quad (5.9)$$

wobei $\hat{\beta}_{ML} = \hat{\beta}_{KQ}$. Im Gegensatz zu $\hat{\sigma}_{KQ}^2$ ist $\hat{\sigma}_{ML}^2$ verzerrt. Allerdings ist er umso weniger verzerrt, je größer n bei gleichbleibender Anzahl an Variablen K .

Auch hier sei für eine ausführliche Herleitung auf die Literatur verwiesen, z. B. Hübler (1989) und Judge et al. (1998).

5.2.7 Das Bestimmtheitsmaß

Das Bestimmtheitsmaß R^2 gibt an, welcher Anteil der Varianz von y durch das Regressionsmodell erklärt wird. Es ist definiert als

$$R^2 = \frac{\frac{1}{n} \sum (\hat{y}_i - \bar{y})^2}{\frac{1}{n} \sum (y_i - \bar{y})^2} = 1 - \frac{\frac{1}{n} \sum (\hat{\epsilon}_i - \bar{\hat{\epsilon}})^2}{\frac{1}{n} \sum (y_i - \bar{y})^2}. \quad (5.10)$$

Der zweite Zähler ist dabei die mittlere quadratische Abweichung der geschätzten Residuen $\hat{\epsilon}_i = y_i - x_i' \hat{\beta}$.

Es gilt immer:

$$0 \leq R^2 \leq 1.$$

Je näher R^2 an 1 ist, desto besser ist der lineare Zusammenhang zwischen y und X .

Je mehr Variablen man zum Regressionsmodell hinzufügt, desto mehr steigt R^2 . Das heißt aber nicht, dass das Modell an sich besser wird. Das korrigierte Bestimmtheitsmaß ist so modifiziert, dass es nicht prinzipiell zunimmt, wenn eine weitere Variable hinzugefügt wird. Es ist wie folgt definiert:

$$\bar{R}^2 = 1 - \frac{n-1}{n-K} (1 - R^2). \quad (5.11)$$

\bar{R}^2 ist immer kleiner als R^2 , so lange $K \geq 1$ (vgl. Hübler (1989)).

5.2.8 Robuste Schätzungen

Die Robustheit einer Schätzung bezieht sich auf ihre Fähigkeit, Schätzer zu liefern, die unsensibel gegenüber Fehlspezifikationen des Modells sind. Unter den Annahmen des allgemeinen linearen Modells ist der KQ-Schätzer der beste unverzernte Schätzer. Sind die Fehler dagegen nicht normalverteilt, ist es häufig möglich, effizientere Schätzer als den KQ-Schätzer zu finden (vgl. Judge et al. (1998)). In dieser Arbeit werden robuste Standardfehler berechnet.

5.2.9 Interpretation der Koeffizienten

Im linearen Grundmodell nach Gleichung 5.1 entsprechen die Koeffizienten den partiellen Ableitungen:

$$\frac{\partial y}{\partial x_k} = \beta_k, k = 1, \dots, K.$$

β_k wird als marginale Größe bezeichnet und gibt den absoluten Preis für eine Einheit der Variable x_k an, d. h. erhöht sich x_k um eine Einheit, erhöht sich y , in dieser Arbeit der Preis, um β_k (vgl. Hübler (1989)). Man betrachtet also absolute Preissteigerungen, unabhängig vom Preisniveau.

Beim log-log-Modell wird β_k als Elastizität interpretiert, da

$$\frac{\partial \ln y}{\partial \ln x_k} = \frac{\frac{dy}{y}}{\frac{dx_k}{x_k}} = \epsilon_{xy} = \beta_k$$

(vgl. Hübler (1989).)

“Die Elastizität bezeichnet das Verhältnis relativer Änderungen von x und y , bezogen auf die jeweiligen absoluten Größen von x bzw. y ” (Woll; 2000).

$$\epsilon_{xy} = \frac{\beta_k \text{-\%ige Änderung in } y}{1 \text{-\%ige Änderung in } x_k},$$

d. h. erhöht sich x_k um 1 %, erhöht sich y um β_k %.

Im log-linearen Modell gibt β_k an, um wieviel Prozent sich y verändert, wenn sich x_k um eine Einheit erhöht:

$$\begin{aligned} \ln y &= \beta_k x_k \\ \frac{\partial \ln y}{\partial x_k} &= \beta_k \\ \frac{dy}{y} &= \beta_k dx_k \end{aligned}$$

Diese Aussage ist unabhängig vom Ausgangsniveau von x_k . Soll dieses in die Betrachtung mit einbezogen werden, so ist der o. g. Ausdruck um x_k zu erweitern:

$$\begin{aligned} \frac{dy}{y} x_k &= \beta_k (dx_k) x_k \\ \frac{\frac{dy}{y}}{\frac{dx_k}{x_k}} &= \beta_k x_k \end{aligned} \tag{5.12}$$

$\beta_k x_k$ gibt an, um wieviel Prozent y an der Stelle x_k steigt, wenn x_k um 1 % steigt.

Sind die Variablen unstetig, wie das bei Dummy-Variablen der Fall ist, so gibt β_k an, um wieviel der Wert von y im Fall $D = 1$ vom Fall $D = 0$ abweicht.

Wird in Gleichung 5.5 y durch $\ln y$ ersetzt, gibt δ an, um wieviel Prozent y steigt, wenn sich $D = 0$ auf $D = 1$ ändert. Es gilt

$$\begin{aligned} \ln(y|D = 1) - \ln(y|D = 0) &= \beta_1 + \delta - \beta_1 \\ \ln\left(\frac{(y|D = 1)}{(y|D = 0)}\right) &= \delta \\ \frac{(y|D = 1)}{(y|D = 0)} &= \exp\{\delta\} \\ \frac{(y|D = 1) - (y|D = 0)}{(y|D = 0)} &= \exp\{\delta\} - 1 \\ &\cong \delta. \end{aligned}$$

Wie bereits in Abschnitt 5.2.2 beschrieben, müssen bei m möglichen Zuständen nur $m - 1$ Zustände modelliert werden. Der nicht modellierte Zustand ist der Referenzzustand. Der Koeffizient kann somit auch als prozentualer Auf- bzw. Abschlag gegenüber dem Referenzzustand interpretiert werden.

Ist die abhängige Variable logarithmiert, geben β_k bzw. δ immer einen relativen Aufschlag an. Die absolute Preisänderung ist also abhängig vom Preisniveau.

Kapitel 6

Deskriptive Statistiken

Dieses Kapitel beginnt mit der Analyse einzelner Variablen. Es werden deskriptive Statistiken vorgestellt, die helfen sollen, die Häuser und die einzelnen Bezirke besser kennenzulernen und zu vergleichen.

6.1 Deskriptive Statistiken zu den Häusern

Die Verteilung der Beobachtungen auf Bezirke und Jahre kann dem Anhang B entnommen werden. In fünf Bezirken wurde im Betrachtungszeitraum kein einziges Haus verkauft, nämlich in Mitte, Tiergarten, Friedrichshain, Kreuzberg und Schöneberg. Mit knapp 20 % finden sich die meisten Beobachtungen in Reinickendorf. Betrachtet man die Jahre, sieht man, dass die Beobachtungen sich auch nicht gleichmäßig über die Jahre verteilen. Mit 16,10 % wurden im Jahr 1999 die meisten Häuser verkauft und im Jahr 1991 mit 7,05 % die wenigsten.

Das durchschnittliche Haus war 40,57 Jahre alt, stand auf einem 585,94 m^2 großen Grundstück, hatte eine Geschossfläche von 146,30 m^2 und kostete DM 600.125 (s. Tabelle 6.1 und 6.2 auf Seite 40). Ein Alter von Null Jahren deutet auf einen Neubau hin. Mehr als 5 % aller Häuser waren Neubauten. Ist die angegebene Anzahl an Beobachtungen geringer als 11.503, so bedeutet dies, dass die Variable nicht bei allen Beobachtungen erfasst wurde.

Welche Preise in den einzelnen Bezirken erzielt wurden, lässt sich Tabelle 6.2 auf Seite 40 entnehmen. Mit Abstand war Wilmersdorf der teuerste Bezirk. Hier kostete ein Haus im Durchschnitt DM 1.894.346. Danach folgten Zehlendorf und Charlottenburg mit einem Durchschnittspreis von DM 1.095.233 bzw. DM 1.092.321. Die geringsten Preise wurden in Treptow erzielt. Hier

Tabelle 6.1: *Deskriptive Statistiken für Alter, Grundfläche und Geschossfläche*

Variable	Obs	Mean	Median	Std. Dev.	Min	Max
Grundstücksfläche	11.485	585,94	530,00	334,0006	70	7045
Geschossfläche	11.500	146,30	134,00	64,9442	24	2885
Alter	11.484	40,57	39,00	25,5335	0	200

lag der Preisdurchschnitt bei DM 311.638. Es folgten Weißensee und Hohenschönhausen mit DM 360.976 bzw. DM 384.866. Das teuerste Haus wurde in Zehlendorf für DM 8.200.000 verkauft, das billigste für nur DM 17.500 in Marzahn.

Tabelle 6.2: *Preise in DM in den einzelnen Bezirken*

Bezirk	Obs	Mean	Median	Std. Dev.	Min	Max
Wedding	12	444.500	457.500	55.505	350.000	504.000
Prenzl. Berg	7	588.677	490.000	256.997	430.000	1.130.737
Charlottenb.	85	1.092.321	980.000	457.346	350.000	2.400.000
Spandau	1.275	526.573	478.000	268.630	110.000	2.400.000
Wilmerdorf	150	1.894.346	1.550.000	1.292.409	120.000	7.500.000
Zehlendorf	1.177	1.095.233	885.000	690.289	265.000	8.200.000
Steglitz	775	721.489	635.000	382.184	160.000	4.000.000
Tempelhof	1.366	558.180	535.000	178.511	93.380	1.750.000
Neukölln	1.091	503.148	490.000	172.716	120.000	1.650.000
Treptow	352	311.638	300.000	126.573	35.157	850.000
Köpenick	454	441.455	400.000	246.891	40.000	2.500.000
Lichtenberg	234	396.966	352.500	287.281	100.000	2.800.000
Weißensee	259	360.976	350.000	131.466	92.880	1.100.000
Pankow	740	444.931	447.250	172.767	30.000	2.100.000
Reinickendorf	2.227	607.566	560.000	271.492	120.000	2.775.000
Marzahn	382	391.066	360.000	193.052	17.500	2.120.000
Hohenschönh.	337	384.866	383.151	153.936	20.000	1.520.000
Hellersdorf	580	392.873	376.160	170.089	46.500	2.500.000
Total	11.503	600.125	500.000	429.543	17.500	8.200.000

Die Preise in den einzelnen Bezirken sind in Abbildung 6.1 noch einmal veranschaulicht. Man sieht, dass es sich bei den fünf Bezirken ohne Beobachtungen um die Citybezirke handelt (für die Namen der Bezirke s. Abbildung 4.1 auf Seite 22). An den dunkleren Flächen im Westen erkennt man tendenziell höhere Preise.

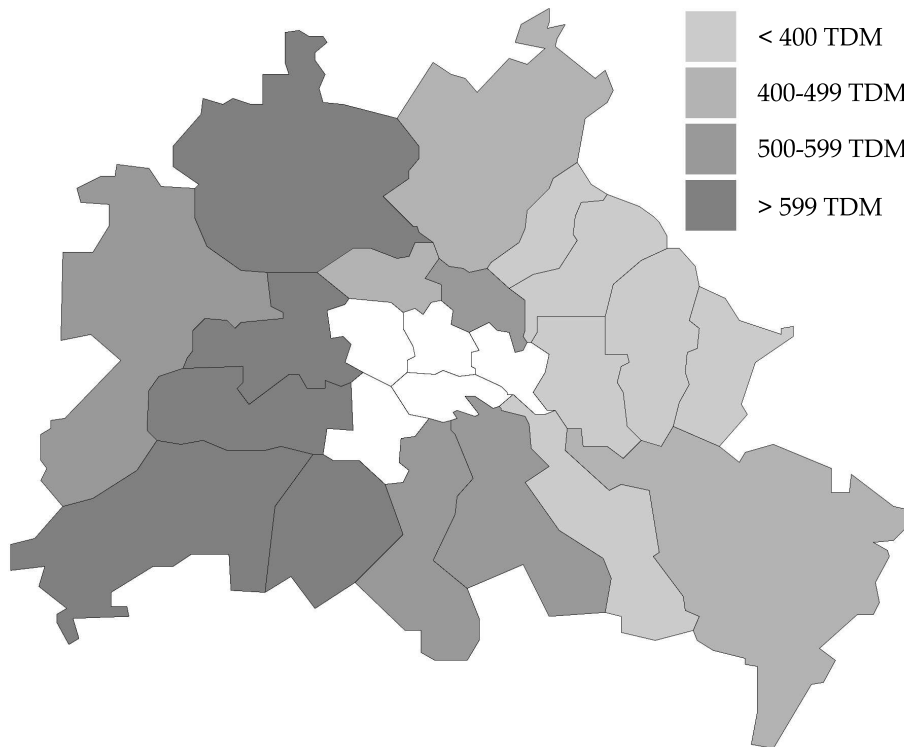


Abbildung 6.1: Preisniveau in Berlin

Das wird auch in Tabelle 6.3 auf Seite 42 bestätigt. Das Preisniveau in den westlichen Bezirken war um das 1,7fache höher als in den östlichen. Ein Haus kostete im Westteil der Stadt durchschnittlich DM 682.326, während es im Ostteil nur knapp DM 400.000 kostete. Im Laufe der Zeit haben sich die Preise in den beiden Stadthälften angeglichen. So betrug der Preis im Westteil der Stadt 1991 noch das 2,3fache des Preises im Ostteil, im Jahr 2000 dagegen nur noch das 1,4fache. Der Minimalpreis war im Westteil mehr als fünfmal so hoch wie im Ostteil, der Maximalpreis immer noch fast dreimal so hoch. Ebenso wurden im Westteil mehr als doppelt so viele Häuser verkauft wie im Ostteil der Stadt.

Tabelle 6.3: *Deskriptive Statistiken für den Preis, getrennt nach Ost und West*

Variable	Obs	Mean	Median	Std. Dev.	Min	Max
Ost	3.345	399.648	380.000	192.640	17.500	2.800.000
West	8.158	682.326	570.000	470.869	93.380	8.200.000

Die Entwicklung des durchschnittlichen Preises über die Jahre ist in Abbildung 6.2 getrennt für Ost-, West- und Gesamt-Berlin geplottet. Auch hier zeigt sich die Angleichung der Preise, allerdings auf einem insgesamt niedrigeren Niveau.

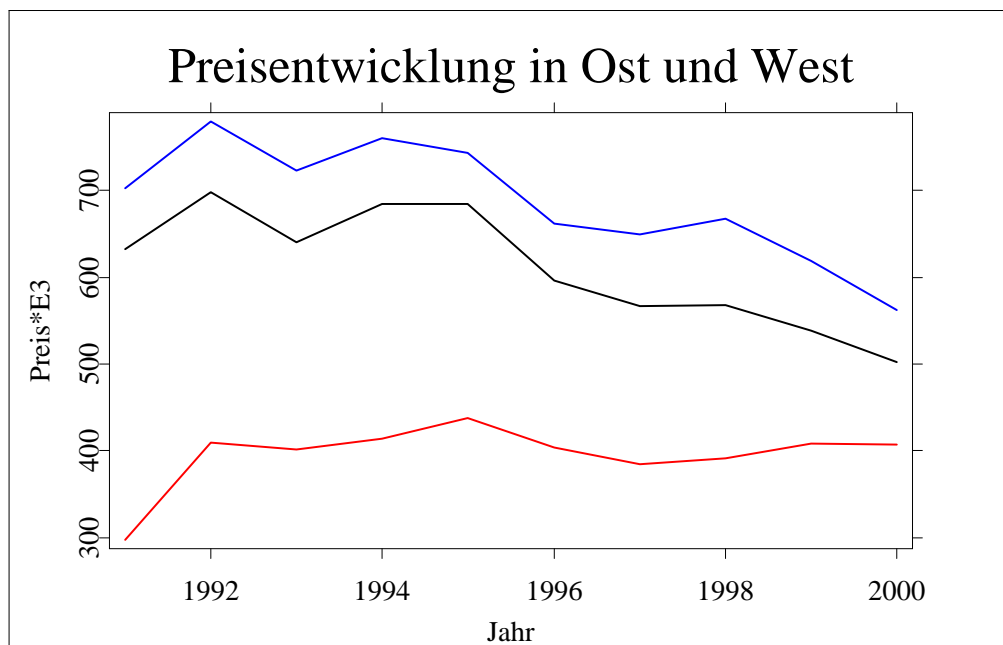


Abbildung 6.2: Durchschnittliche Preise in Ost-Berlin (rot), West-Berlin (blau) und Gesamt-Berlin (schwarz) über die Jahre 1991 - 2000

In Abschnitt 5.2.2 wurde bereits erläutert, dass bei kategoriellen Variablen ein Zustand als Referenzzustand genommen wird. Wie sich in Abschnitt 8.1 zeigen wird, werden die üblichen Zustände als Referenzzustände der Regression benutzt. Am üblichsten ist ein bezugsfreies Einzelhaus mit einem Vollgeschoss und Satteldach, in normalem Zustand, von privat an privat verkauft. Tabelle 6.4 auf Seite 43 zeigt den Anteil an Häusern an bestimmten Kategorien.

Tabelle 6.4: Anteil der Häuser an bestimmten Kategorien

Dachform			
Flachdach	11,41 %	Pultdach	3,86 %
Satteldach	61,13 %	Berliner Dach	0,06 %
Walmdach	17,76 %	Mansardendach	3,33 %
Zeltdach	2,22 %	sonstige	0,23 %
Erwerbergruppe			
Privat	96,58 %	GbR	1,23 %
sonst. jur. Person	2,09 %	sonstige	0,10 %
Gebäudestellung			
Einzelhaus	55,33 %	Doppelhaushälfte	22,33 %
Reihenhaus	16,79 %	Reihenendhaus	5,55 %
Geschossanzahl			
eins	57,37 %	zwei	41,03 %
drei oder vier	1,60 %		
Veräußerergruppe			
Privat	84,39 %	GbR	0,47 %
BRD	1,18 %	Land Berlin	0,30 %
sonst. öffentl. Hand	0,05 %	gemeinn. Wohnunt.	2,98 %
sonst. jur. Person	10,24 %	Kirche	0,06 %
Landeseigene Wohn- ungsunternehmen	0,26 %	sonstige	0,07 %
Verfügbarkeit			
bezugsfrei	91,26 %	teilvermietet	0,26 %
vermietet	1,40 %	Mieterkauf	2,84 %
Nutzungsrecht	0,12 %	Leerstand	0,43 %
nicht bezugsfertig	3,69 %		
Zustandsnote			
gut	29,82 %	normal	61,47 %
schlecht	8,71 %		
<i>Anmerkung:</i> Die Variablen Dachform , Anzahl der Vollgeschosse und Zustandsnote wiesen missings auf. Die Anteile beziehen sich auf die Beobachtungen, bei denen das Merkmal erfasst war.			

Wie viele Häuser besondere Eigenschaften aufwiesen, zeigt Tabelle 6.5 auf Seite 44.

Tabelle 6.5: *Anteil der Häuser mit besonderen Eigenschaften*

Eigenschaft	Anteil
Dachgeschossausbau	52,14 %
Geschäftliche Bindung	0,09 %
Persönliche Verhältnisse	2,98 %
Ungewöhnl. Verhältnisse Grundstück	2,50 %
Ungewöhnl. Verhältnisse Geschäftsverkehr	1,97 %
Wasserlage	0,72 %

6.2 Deskriptive Statistiken zu den Bezirken

Die deskriptiven Statistiken für Gesamt-Berlin finden sich in Tabelle 6.6, die für die einzelnen Bezirke in den Tabellen 6.7 auf Seite 45 und 6.8 auf Seite 46.

Tabelle 6.6: *Deskriptive Statistiken für Berlin*

Fläche	89.171
Anteil Grünfläche	40,72 %
Anteil Waldfläche	17,58 %
Anteil Wasserfläche	6,52 %
Anteil bebaute Fläche	41,12 %
Anteil Kinder	14,71 %
Anteil Senioren	13,90 %
Ausländeranteil	12,21 %
Einwohner/ha	38,61
Alter	40,92

Die Fläche Berlins beträgt 89.171 ha. Der größte Bezirk ist mit einer Fläche von 12.776 ha Köpenick, der kleinste Friedrichshain mit 978 ha. Wie aus Tabelle 6.7 auf Seite 45 ersichtlich, hat mit 76,69 % Köpenick relativ die meisten Grünflächen, wobei diese sich aus den Erholungsflächen, Landwirtschaftsflächen, Waldflächen und Wasserflächen zusammensetzen. Erholungsflächen sind "unbebaute Flächen, die dem Sport und der Erholung dienen" (Statistisches Landesamt Berlin; 2000).

Tabelle 6.7: Anteil verschiedener Teilflächen an der Gesamtfläche des Bezirks im Jahr 2000

Bezirk	Fläche	Anteil Grün- fläche	Anteil Wald- fläche	Anteil Wasser- fläche	Anteil bebaute Fläche
Mitte	1.069	11,23 %	0,00 %	3,65 %	55,94 %
Tiergarten	1.341	28,33 %	0,00 %	6,56 %	51,68 %
Wedding	1.537	20,75 %	0,00 %	0,96 %	52,37 %
Prenzl. Berg	1.095	14,43 %	0,00 %	0,00 %	54,43 %
Friedrichshain	978	20,35 %	0,00 %	9,51 %	49,28 %
Kreuzberg	1.038	9,25 %	0,00 %	2,22 %	59,34 %
Charlottenburg	3.033	25,35 %	2,37 %	3,23 %	46,06 %
Spandau	9.191	46,10 %	18,00 %	9,56 %	39,54 %
Wilmerdorf	3.439	54,87 %	44,26 %	5,26 %	28,53 %
Zehlendorf	7.053	56,87 %	34,55 %	15,11 %	31,41 %
Schöneberg	1.229	13,43 %	0,00 %	0,16 %	52,07 %
Steglitz	3.197	13,67 %	0,13 %	1,88 %	66,31 %
Tempelhof	4.080	14,04 %	1,15 %	0,98 %	59,58 %
Neukölln	4.493	23,46 %	0,07 %	1,56 %	56,69 %
Treptow	4.066	31,87 %	4,82 %	0,96 %	42,72 %
Köpenick	12.776	76,69 %	51,82 %	16,63 %	15,62 %
Lichtenberg	2.630	24,56 %	0,68 %	2,55 %	52,17 %
Weißensee	3.016	42,74 %	0,56 %	1,19 %	39,99 %
Pankow	6.190	52,39 %	19,53 %	1,74 %	28,26 %
Reinickendorf	8.946	42,25 %	21,59 %	8,18 %	39,82 %
Marzahn	3.196	20,96 %	2,63 %	1,78 %	51,85 %
Hohenschönh.	2.599	33,97 %	1,04 %	1,77 %	43,32 %
Hellersdorf	2.979	27,93 %	2,26 %	1,91 %	52,67 %
Mean	3.877	30,67 %	8,89 %	4,23 %	46,65 %
Median	3.033	25,35 %	1,04 %	1,91 %	51,68 %
Std. Dev.	3.064	0,1756	0,1539	0,0463	0,1200
Min	978	9,25 %	0,00 %	0,00 %	15,62 %
Max	12.776	76,69 %	51,86 %	16,63 %	66,31 %

Im Westteil der Stadt hat Zehlendorf mit 56,87 % die meisten Grünflächen, Kreuzberg mit nur 9,25 % die wenigsten. Betrachtet man nur die Bezirke, in denen Häuser verkauft wurden, bleibt Köpenick der grünste Bezirk. Der am wenigsten grüne Bezirk ist dann Steglitz, dessen Fläche zu 13,67 % begrünt ist. 40,72 % Berlins sind mit Grünflächen bedeckt (s. Tabelle 6.6 auf Seite 44). Köpenick hat mit 51,82 % und 16,63 % die größten Wald- bzw. Wasserflächen. Hier liegt der Müggelsee, die größte Wasserfläche Berlins. Danach folgen Wilmerdorf, wo es mit dem Grunewald die zweitgrößte Waldfläche

gibt, bzw. Zehlendorf, wo der Wannsee liegt.

Tabelle 6.7 kann ebenfalls entnommen werden, dass Steglitz der am stärksten bebaute Bezirk ist. Dessen Fläche ist zu 66,31 % bebaut, dicht gefolgt von Tempelhof (59,58 %) und Kreuzberg (59,34 %). Insgesamt ist die Fläche Berlins zu 17,58 % mit Wald und zu 6,52 % mit Wasser bedeckt sowie zu 41,12 % bebaut (s. Tabelle 6.6 auf Seite 44). Die bebauten Flächen entsprechen den vom Statistischen Landesamt Berlin ausgewiesenen Gebäude- und Freiflächen. Dabei handelt es sich um "Flächen mit Gebäuden (Gebäudeflächen) und unbebaute Flächen (Hofflächen), die Zwecken der Gebäude untergeordnet sind (z. B. Vor- und Hausgärten, Spielplätze, Stellplätze usw.)" (Statistisches Landesamt Berlin; 2000). Außer aus Grünflächen und bebauten Flächen besteht die Bezirksfläche noch aus Betriebs- und Verkehrsflächen sowie aus Flächen anderer Nutzung.

Köpenick war außerdem der am dünnsten besiedelte Bezirk. Hier lebten im Betrachtungszeitraum im Mittel nur 8,72 Einwohner pro Hektar Bezirksfläche. Kreuzberg war dagegen mit 147,50 Einwohnern/ha der am stärksten besiedelte Bezirk. Durchschnittlich hatte Berlin 38,61 Einwohner/ha. Die jüngsten Einwohner hatte Hellersdorf. Das mittlere Alter war mit nur 32,70 Jahren deutlich geringer als der Gesamt-Berliner Durchschnitt von 40,92 Jahren. 24,97 % der Bevölkerung waren in Hellersdorf jünger als 15 Jahre und nur 6,78 % 65 und älter. Damit war Hellersdorf der Bezirk mit den meisten Kindern (Bevölkerung unter 15 Jahren) und den wenigsten Senioren (Bevölkerung im Alter von 65 Jahren und älter) (s. Tabelle 6.8).

Tabelle 6.8: *Deskriptive Statistiken zu den Bezirken*

Bezirk	Einwohner pro ha	Alter	Anteil Kinder	Anteil Senioren	Anteil Ausländer
Mitte	73,18	39,63	14,30 %	11,26 %	9,63 %
Tiergarten	68,32	39,49	13,76 %	11,98 %	26,25 %
Wedding	107,56	39,15	16,11 %	12,82 %	27,92 %
Prenzl. Berg	129,40	39,22	13,06 %	11,76 %	5,66 %
Friedrichshain	107,23	39,59	13,49 %	12,65 %	6,20 %
Kreuzberg	147,50	35,86	17,65 %	8,30 %	33,00 %
Charlottenburg	59,49	43,49	11,52 %	16,16 %	18,09 %
Spandau	24,35	42,70	14,46 %	16,22 %	10,98 %
Wilmerdorf	41,65	45,26	10,69 %	18,51 %	13,09 %
Zehlendorf	13,93	44,81	13,68 %	18,78 %	9,72 %
Schöneberg	124,30	40,27	13,47 %	12,53 %	21,73 %

Fortsetzung auf der nächsten Seite

Tabelle 6.8: *Deskriptive Statistiken zu den Bezirken - Fortsetzung*

Bezirk	Einwohner pro ha	Alter	Anteil Kinder	Anteil Senioren	Anteil Ausländer
Steglitz	59,94	44,10	12,85 %	17,91 %	9,39 %
Tempelhof	46,64	44,06	13,32 %	17,81 %	9,31 %
Neukölln	69,18	40,77	14,96 %	13,95 %	19,32 %
Treptow	26,86	42,19	14,03 %	15,94 %	3,34 %
Köpenick	8,72	43,74	13,38 %	16,61 %	3,26 %
Lichtenberg	62,14	41,92	12,66 %	12,68 %	8,29 %
Weißensee	20,32	42,19	14,06 %	14,95 %	2,94 %
Pankow	18,48	41,69	14,29 %	14,62 %	3,47 %
Reinickendorf	27,93	43,43	14,34 %	16,91 %	9,19 %
Marzahn	48,97	35,53	18,67 %	7,10 %	3,29 %
Hohenschönh.	44,79	34,31	19,48 %	7,17 %	4,50 %
Hellersdorf	45,41	32,70	24,97 %	6,78 %	2,31 %
Mean	59,84	40,73	14,75 %	13,63 %	11,34 %
Median	54,23	41,69	14,03 %	13,95 %	9,31 %
Std. Dev.	39,2889	3,4526	0,0303	0,0371	0,0886
Min	8,72	32,70	10,69 %	6,78 %	2,31 %
Max	147,50	45,26	24,97 %	18,78 %	33,00 %

Der Bezirk mit den ältesten Einwohnern war Wilmersdorf. Hier waren die Einwohner durchschnittlich 45,26 Jahre alt und lebten mit 10,69 % die wenigsten Kinder. Mit 18,78 % gab es allerdings die meisten Senioren in Zehlendorf. Über die Jahre 1991 bis 2000 stieg das Durchschnittsalter in Berlin von 40,50 Jahren auf 41,76 Jahre. Der durchschnittliche Anteil an Kindern fiel von 15,85 % im Jahr 1991 auf 13,44 % im Jahr 2000 und betrug im Durchschnitt 14,71 %. Umgekehrt stieg der Anteil an Senioren von 14,00 % im Jahr 1991 auf 14,61 % im Jahr 2000 und betrug im Mittel 13,90 %.

In Kreuzberg lebten im Betrachtungszeitraum mit 33,00 % relativ zur Bevölkerung die meisten Ausländer, in Hellersdorf dagegen nur 2,31 %. Der durchschnittliche Ausländeranteil Berlins betrug im Betrachtungszeitraum 12,21 %. Er stieg von 10,31 % im Jahr 1991 auf 12,87 % im Jahr 2000.

Die Anzahl der U- und S-Bahnhöfe in den einzelnen Bezirken kann man in Abbildung 6.3 auf Seite 48 ablesen. In den Außenbezirken verkehrt vor allem die S-Bahn, während das U-Bahnnetz im Zentrum und im Westteil der Stadt dichter ist. Das Verkehrsnetz ist in den vergangenen Jahren immer weiter ausgebaut worden. Insgesamt hatte Berlin Ende 2000: 120 S-Bahnstationen und 165 U-Bahnstationen.

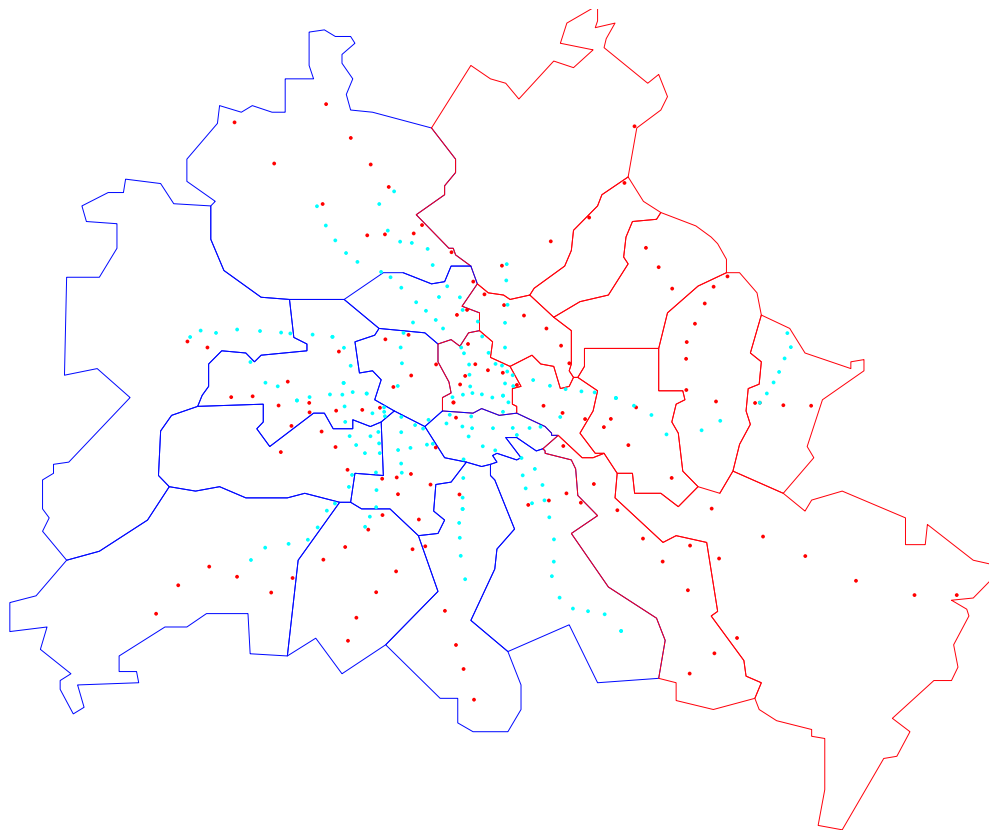


Abbildung 6.3: Anzahl der S-Bahnstationen (rot) und U-Bahnstationen (blau) in den Bezirken im Jahr 2000

Kapitel 7

Missings

Es wird zunächst beschrieben, bei welchen Variablen missings auftauchen und wie deren zeitliche Struktur ist. Da alle Beobachtungen, bei denen die Ausprägung einer Variablen missing ist, aus der Schätzung der Regressionsfunktion herausfallen, werden für zwei wichtige Variablen, das Haushaltsnettoeinkommen und die Arbeitslosigkeit, die fehlenden Werte geschätzt, bevor im nächsten Kapitel zur Regression übergegangen wird.

7.1 Struktur der missings

Bei der Auswahl der Variablen muss beachtet werden, dass Beobachtungen, bei denen eine der Variablen die Ausprägung missing aufweist, aus der Regression ausgeschlossen werden. Dies führt dazu, dass Variablen, die bei vielen Beobachtungen die Ausprägung missing aufweisen, nicht für die Regression benutzt werden können, damit der Datensatz möglichst groß bleibt. Zunächst wurde also für die Hausvariablen untersucht, welche Variablen besonders viele missings aufwiesen. Variablen, bei denen mehr als 10 % missing waren, wurden für die Regression nicht in Betracht gezogen. Bei einem Gesamtumfang von 11.503 Beobachtungen bedeutet das, dass das Merkmal bei maximal 1.150 Beobachtungen missing sein durfte. Variablen wie **tatsächliche Bau-massenzahl** (Anzahl missings: 4.703), **Wohnfläche** (10.413), **Objektförderung** (11.461) und **Keller** (4.376) wurden daher nicht benutzt. Aufgrund der missings in den benutzten Variablen wurden für die Regression nur 10.136 Beobachtungen verwendet.

Die Anzahl der missings in den letztendlich benutzten Hausvariablen ist in Tabelle 7.1 auf Seite 50 dargestellt.

Tabelle 7.1: *Anzahl missings für Hausvariablen*

Variable	Anzahl
Alter	19
Grundstücksfläche	18
Geschossfläche	3
Blocklage	66
Zustand	208
Geschossanzahl	961
<i>Anmerkung:</i> Die hier nicht aufgeführten Variablen wiesen keine missings auf.	

Besonders häufig waren diejenigen Variablen nicht erfasst, die die Lage des Hauses beschreiben. Hierzu gehörten die **Zulässige Geschossflächenzahl** (3.157), die **Typische Geschossflächenzahl** (2.013), die **Stadträumliche Wohnlage** (4.019), **Stadterneuerung** (11.471), die **Typische Nutzungsart** (1.790) und die **Zulässige Art der Nutzung** (5.517). Durch diese Variablen ist es möglich, die Lage eines Objekts genauer zu beschreiben, als dies allein auf Bezirksebene möglich ist.

Die Geschossflächenzahl (GFZ) berechnet sich nach §20 BauNVO aus

$$\frac{\text{Geschossfläche}}{\text{Grundstücksfläche}}.$$

Die GFZ ist somit ein Maß baulicher Dichte. Die zulässige GFZ wird im Bebauungsplan festgelegt.

Die Baumassenzahl (BMZ) gibt nach §21 BauNVO an, "wieviel Kubikmeter Baumasse je Quadratmeter Grundstücksfläche [...] zulässig sind":

$$\text{Grundstücksfläche} \cdot \text{BMZ} = \text{Kubikmeter Baumasse}$$

Die Baumassenzahl ist die Richtzahl für die Bebauung in Industriegebieten, "da hier nicht wie bei Wohngebäuden mit normalen Geschosshöhen, sondern mit m^3 umbauten Raumes gerechnet werden muss" (Gerhards und Keller; 2002). Diese Variable soll daher in dieser Arbeit nicht weiter betrachtet werden.

§17 BauNVO legt Obergrenzen für die GFZ und die Baumassenzahl vor. Sie unterscheidet dabei nach Siedlungsgebieten, die in den §§1 - 10 BauNVO näher definiert werden. Das zulässige Maß baulicher Nutzung findet sich im

Anhang A Tabelle A.2 auf Seite 99.

Die stadträumliche Wohnlage gibt an, ob es sich um eine einfache, mittlere, gute oder sehr gute Lage handelt. Dieser Begriff wird im Mietspiegel verwendet. Der Mietspiegel "für freifinanzierten Wohnraum dient unter anderem als Richtlinie zur Ermittlung der ortsüblichen Vergleichsmieten" (Gerhards und Keller; 2002). Für die Charakterisierung der einzelnen Wohnlagen s. Anhang A Tabelle A.2.

Wohngebiete können vom produzierenden Gewerbe oder Dienstleistungsgewerbe genutzt werden sowie einen mischgebietstypischen Charakter oder reine Wohnbebauung aufweisen. §1 Abs. 2 BauNVO unterscheidet u. a. folgende Arten baulicher Nutzung: Gewerbegebiet, Dorfgebiet, Mischgebiet, Sondergebiet, allgemeines Wohngebiet, reines Wohngebiet und Kleinsiedlungsgebiet. Die für diese Arbeit wichtigsten Gebiete, nämlich das Wohngebiet, das Mischgebiet und das Gewerbegebiet, werden im Anhang A Tabelle A.2 näher beschrieben.

Betrachtet man die zeitliche Struktur der missings, so fällt auf, dass manche Variablen erst ab einem bestimmten Zeitpunkt erfasst wurden. So fehlte die **Tatsächliche Baumassenzahl** bei 90,84 % der Beobachtungen bis 1995, aber nur bei 4,36 % der Beobachtungen ab 1996, und dort vor allem in den Bezirken Wilmersdorf und Treptow. Auch die Variable **Keller** wurde erst ab 1996 regelmäßig erfasst. Hier betrug die Fehlquote 86,31 % für den Zeitraum bis 1995 und 2,75 % für den Zeitraum ab 1996. Wieder sind Wilmersdorf und Treptow die Bezirke, in denen auch nach 1996 die Variable gar nicht bzw. so gut wie gar nicht erfasst wurde.

Gerade bei der Variable **Keller** stellt sich die Frage, ob ein fehlender Eintrag tatsächlich auf eine fehlende Unterkellerung hinweist. Von den Häusern, bei denen die Variable erfasst wurde, waren 9,2 % nicht unterkellert, 75,8 % hatten einen Keller und 15,0 % einen Teilkeller. Bedenkt man die zusätzlichen Kosten, die durch eine Unterkellerung entstehen, ist die Annahme, dass fehlende Erfassung auf einen fehlenden Keller hinweist, plausibel. Da die Annahme trotz Plausibilität spekulativ ist, wird die Variable **Keller** nicht weiter betrachtet.

Die **Zulässige Geschossflächenzahl** wurde in den Bezirken Zehlendorf, Treptow, Köpenick, Pankow, Marzahn, Hohenschönhausen und Hellersdorf über den gesamten Zeitraum schlecht erfasst. Die **Typische Geschossflächenzahl** fehlte 1991 und 1992 in 99,06 % der Fälle, danach nur noch in

3,27 %. Nur in Hellersdorf wurde sie auch danach nur unregelmäßig erfasst. Die **Wohnlage** wurde erst ab 1995 erfasst. Davor fehlte sie in 98,65 % der Beobachtungen, ab 1995 in 2,1 %. Auch die **Typische Nutzungsart** wurde offensichtlich erst nach 1993 erfasst. Bis dahin fehlte sie in 99,82 % der Fälle, danach in 0,86 %. Bei der Variable **Zulässige Art der Nutzung** wiesen bis einschließlich 1995: 78,86 % der Beobachtungen die Ausprägung missing auf, ab 1996 immerhin noch 16,09 %. Vor allem in den Ost-Berliner Bezirken wurde sie auch nach 1995 nur unregelmäßig erfasst.

Es zeigte sich also, dass die Variablen **Typische Geschossflächenzahl**, **Wohnlage** und **Typische Nutzungsart** ab 1996 regelmäßig erfasst wurden und für eine Regression benutzt werden können. Im Abschnitt 8.3 soll daher eine Regression für den Teilzeitraum 1996 bis 2000 vorgestellt werden. Die **Tatsächliche Baumassenzahl** wird nicht weiter beachtet, da es sich dabei, wie erwähnt, um eine Richtzahl in Industriegebieten handelt. Auch die Variable **Keller** wird nicht in die Teilregression aufgenommen, da die missings gehäuft in den Bezirken Wilmersdorf und Treptow auftreten und verhindert werden soll, dass diese Bezirke ganz aus der Regression herausfallen.

Die Zuordnung der Häuser zu den Lagekategorien kann Tabelle 7.2 entnommen werden.

Tabelle 7.2: Zuordnung der Häuser zu Lagekategorien für den Zeitraum 1996 - 2000

Typische GFZ			
0	0,06 %	0,1	0,22 %
0,2	1,91 %	0,3	6,05 %
0,4	80,62 %	0,5	0,20 %
0,6	8,13 %	0,7	0,05 %
0,8	0,79 %	0,9	0,05 %
1	1,57 %	1,2	0,20 %
1,5	0,12 %	4	0,03 %
Stadträumliche Wohnlage			
Einfache Lage	23,94 %	Mittlere Lage	51,52 %
Gute Lage	22,60 %	Sehr gute Lage	1,94 %
Typische Nutzungsart			
Wohnbebauung	99,86 %	Sonstige	0,14 %
<i>Anmerkung: Alle Variablen wiesen missings auf. Die Anteile beziehen sich auf die Beobachtungen, bei denen das Merkmal erfasst war.</i>			

Bei der Auswahl der Bezirksvariablen waren die Überlegungen des Abschnitts

4.2 maßgeblich. Im Gegensatz zu missings bei den Hausvariablen können missings bei den Bezirksvariablen geschätzt werden. Beim Haushaltsnettoeinkommen war dies möglich, da ein Trend über die Zeit unterstellt wurde (s. Abschnitt 7.2). Bei der Arbeitslosigkeit wurden Informationen außerhalb des Datensatzes benutzt, um die fehlenden Werte zu schätzen, nämlich die monatlichen Arbeitslosenzahlen (s. Abschnitt 7.3). Da bei den Hausvariablen keine Abhängigkeit von der Zeit unterstellt werden kann und auch keine Informationen außerhalb des Datensatzes vorlagen, hätten fehlende Werte nur mit Hilfe anderer Variablen des Datensatzes geschätzt werden können. Sobald dies aber möglich ist, ist die Annahme der Unabhängigkeit der Variablen verletzt.

7.2 Das Mittlere monatliche Haushaltsnettoeinkommen

Das Mittlere monatliche Haushaltsnettoeinkommen (HHNetto) ist im Jahr 1996 für die Bezirke Mitte und Zehlendorf missing. Das ist nur für Zehlendorf problematisch, da in Mitte im betrachteten Zeitraum kein einziges Einfamilienhaus verkauft wurde. Um zu vermeiden, dass alle Häuser, die im Jahr 1996 in Zehlendorf verkauft wurden, aus der Schätzung herausfallen, muss der fehlende Wert geschätzt werden. Drei verschiedene Vorgehensweisen wurden dabei miteinander verglichen. Ein lineares Modell der Form

$$\text{HHNetto} = \alpha + \beta t + \epsilon$$

ergibt als Wert DM 3.802,94. Aus dem log-linearen Modell der Form

$$\ln(\text{HHNetto}) = \alpha + \beta t + \epsilon$$

ergibt sich als Schätzer für das Mittlere monatliche Haushaltsnettoeinkommen

$$\widehat{\text{HHNetto}} = \exp\{\alpha + \beta t\}$$

DM 3.788,41. Lineare Interpolation ergibt den Wert DM 3.950,00.

Den Wert für 1997 sehe ich als Abweichung vom langfristigen linearen bzw. log-linearen Trend, so dass auch der linear interpolierte Wert so interpretiert werden muss. Dieser Wert wird also verworfen. Bei der Abwägung zwischen dem linearen und log-linearen Modell muss beachtet werden, dass das Statistische Landesamt das Mittlere monatliche Haushaltsnettoeinkommen auf

volle DM 50 rundet, so dass man bei beiden Modellen DM 3.800,00 erhält. Dieser Wert wird in den Datensatz übernommen. In Abbildung 7.1 ist das Haushaltsnettoeinkommen für Zehlendorf gegen die Jahre abgetragen. Die lineare Regressionsgerade ist ebenfalls eingezeichnet.

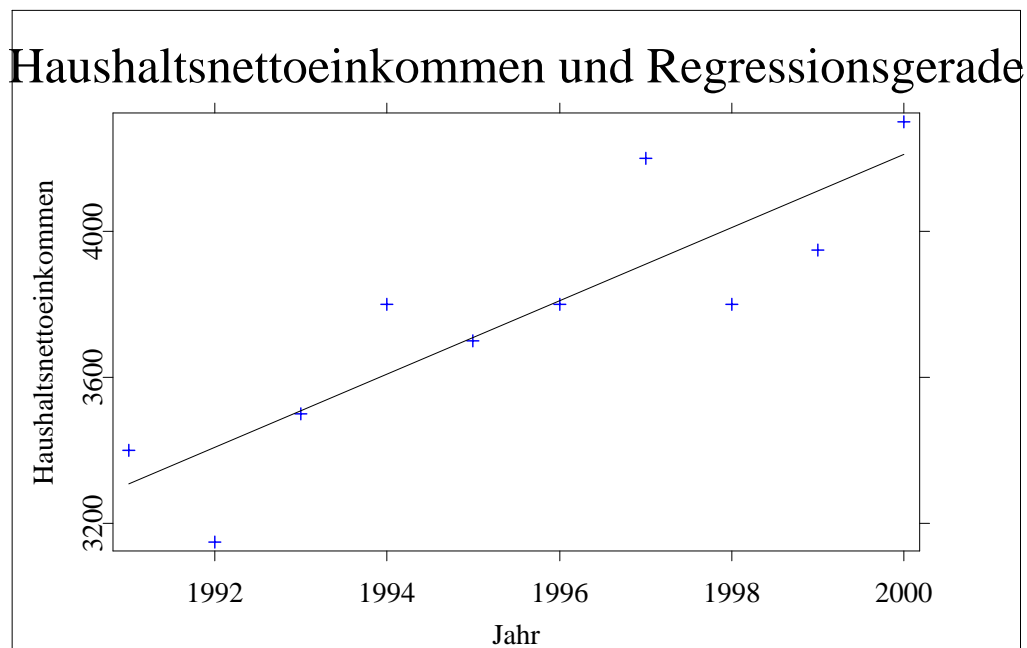


Abbildung 7.1: Das Mittlere monatliche Haushaltsnettoeinkommen in Zehlendorf mit linearer Regressionsgerade

Die deskriptiven Statistiken für das Mittlere monatliche Haushaltsnettoeinkommen finden sich in Tabelle 7.3. Der geringste Wert von DM 1.400 wurde 1991 in Prenzlauer Berg gemessen, der höchste Wert 2000 in Zehlendorf. Im Mittelwert betrug es in Berlin über den gesamten Betrachtungszeitraum DM 2.705,66.

Tabelle 7.3: *Deskriptive Statistiken für das Mittlere monatliche Haushaltsnettoeinkommen*

Variable	Mean	Median	Std. Dev.	Min	Max
HHNetto	2.705,66	2.700	453,90	1.400	4.300

Betrachtet man die einzelnen Bezirke, ist das Mittlere monatliche Haushaltsnettoeinkommen über die Jahre in Friedrichshain am geringsten und in Zehlendorf am höchsten (vgl. Tabelle 7.4 auf Seite 55).

Tabelle 7.4: *Mittleres monatliches Haushaltsnettoeinkommen pro Bezirk*

Bezirk	HHNetto
Mitte	2.400
Tiergarten	2.415
Wedding	2.415
Prenzlauer Berg	2.150
Friedrichshain	2.140
Kreuzberg	2.265
Charlottenburg	2.735
Spandau	2.965
Wilmerdorf	3.120
Zehlendorf	3.760
Schöneberg	2.675
Steglitz	3.120
Tempelhof	2.955
Neukölln	2.565
Treptow	2.635
Köpenick	2.485
Lichtenberg	2.485
Weißensee	2.495
Pankow	2.590
Reinickendorf	3.010
Marzahn	2.945
Hohenschönhausen	2.820
Hellersdorf	3.085

7.3 Arbeitslose

Die Arbeitslosenzahlen in Berlin werden von zwei Stellen erfasst: von den einzelnen Arbeitsämtern und vom Landesarbeitsamt Berlin-Brandenburg. In Berlin gibt es fünf Arbeitsämter: Nord, (zuständig für: Charlottenburg, Wilmerdorf, Pankow, Weißensee, Prenzlauer Berg, Reinickendorf und Spandau), Ost (Lichtenberg, Hohenschönhausen, Marzahn, Hellersdorf), Mitte (Mitte, Friedrichshain, Kreuzberg, Tiergarten, Wedding), Südwest (Tempelhof, Schöneberg, Steglitz, Zehlendorf) und Süd (Neukölln, Treptow, Köpenick). Diese sammeln die Zahlen auf Bezirksebene. Das Landesarbeitsamt Berlin-Brandenburg weist monatliche Zahlen für Gesamt-Berlin aus.

Für die Anzahl der Arbeitslosen auf Bezirksebene gibt es erst seit 1994 vollständige Daten. In den Bezirken Tiergarten, Wedding, Wilmersdorf, Zehlendorf, Schöneberg, Steglitz, Tempelhof, Neukölln und Reinickendorf wurden die Zahlen bereits 1993 vollständig erfasst. Ab Mai 1993 liegen die Werte für Prenzlauer Berg, Kreuzberg, Charlottenburg, Spandau, Weißensee und Pankow vor, ab August 1993 für Mitte, Friedrichshain, Treptow, Köpenick und Lichtenberg. In Marzahn, Hohenschönhausen und Hellersdorf wurden die Daten erst ab September 1993 erfasst. Auf Nachfrage konnte nur das Arbeitsamt Süd für Neukölln die vollständigen Daten für 1992 und den Wert für Dezember 1991 angeben.

Für den gesamten Zeitraum liegen die Zahlen des Landesarbeitsamts Berlin-Brandenburg für Arbeitslose nach Monaten vor, bis 1994 getrennt nach Berlin-West und Berlin-Ost. Allerdings ist der vom Landesarbeitsamt Berlin-Brandenburg erfasste Jahresdurchschnitt an Arbeitslosen in den Jahren 1994 bis 2000 höher als die Summe der Zahlen der einzelnen Arbeitsämter (s. Tabelle 7.5. Die Zahlen der Arbeitsämter sind für 1991 bis 1993, wie erläutert, missing).

Tabelle 7.5: *Arbeitslose im Jahresdurchschnitt*

Jahr	Landesarbeitsamt Berlin-Brandenburg	Arbeitsämter	Abweichung
1991	197.953		
1992	207.058		
1993	203.924		
1994	210.130	208.579	0,7436 %
1995	213.383	212.347	0,4879 %
1996	235.999	235.323	0,2873 %
1997	265.665	264.971	0,2619 %
1998	273.038	272.539	0,1831 %
1999	268.078	267.483	0,2224 %
2000	264.749	264.526	0,0843 %
Mean			0,3244 %

In Tabelle 7.5 ist ebenfalls eine Abweichung ausgewiesen. Diese wurde wie folgt berechnet:

$$\frac{\text{Jahresdurchschnitt Landesarbeitsamt}}{\text{Summe der Jahresdurchschnitte der Arbeitsämter}} - 1.$$

Durchschnittlich weist das Landesarbeitsamt Berlin-Brandenburg eine um

0,3244 % höhere Zahl aus als die Arbeitsämter. Laut Statistischem Landesamt Berlin gibt das Landesarbeitsamt Berlin-Brandenburg als Ursache an, dass aus datenverarbeitungstechnischen Gründen bestimmte Fälle keinem Verwaltungsbezirk zugeordnet werden können und es so zu Differenzen bei der Summenbildung kommen kann. Dies muss bei der Schätzung der Zahlen auf Bezirksebene berücksichtigt werden.

Die Schätzung der Zahlen für 1991 bis 1993 erfolgt in drei Schritten: zuerst müssen die monatlichen Zahlen des Landesarbeitsamtes Berlin-Brandenburg in monatliche Zahlen der Berliner Arbeitsämter umgerechnet werden. Danach wird für jeden Bezirk die Arbeitslosigkeit pro Monat geschätzt, getrennt nach Ost- und West-Berlin. Zum Schluss wird aus diesen monatlichen Arbeitslosenzahlen für jeden Bezirk ein Jahresdurchschnitt berechnet.

Für den ersten Schritt gibt es zwei Möglichkeiten. Einmal könnte man annehmen, dass es sich bei den Abweichungen der Jahresdurchschnitte um zufällige Abweichungen handelt, die über die Jahre keinen Trend aufweisen. Die vom Landesarbeitsamt Berlin-Brandenburg ausgewiesene Anzahl an Arbeitslosen pro Monat müsste also um die mittlere Abweichung von den Zahlen der Arbeitsämter bereinigt werden, um die Anzahl an Arbeitslosen zu berechnen, die die Berliner Arbeitsämter ausgewiesen hätten. Bei der zweiten Möglichkeit nimmt man an, dass die Abweichung von 1994 der beste Schätzer für die Abweichung von 1993 ist. Der Jahresdurchschnitt der Arbeitsämter ergibt sich aus:

$$\text{Jahresdurchschnitt Arbeitsämter} = \frac{\text{Jahresdurchschnitt Landesarbeitsamt}}{1 + \text{Abweichung}}$$

Bei der ersten Möglichkeit würde man also im Nenner 1,003244 einsetzen, bei der zweiten 1,007436 (s. Tabelle 7.5 auf Seite 56). Die Berechnung für die monatlichen Durchschnitte erfolgt analog.

Auch für den zweiten Schritt gibt es mehrere Möglichkeiten der Berechnung. Allen Möglichkeiten gemeinsam ist, dass die Summe der Arbeitslosen des jeweiligen Monats mit dem Anteil des Bezirks multipliziert werden muss, um die Anzahl an Arbeitslosen eines bestimmten Monats in einem bestimmten Bezirk zu schätzen. Die im folgenden geschilderten Möglichkeiten unterscheiden sich darin, wie dieser Anteil berechnet wird.

Die erste Möglichkeit besteht darin, dass man für jeden Monat des Jahres 1994 den Anteil eines jeden Bezirks an der gesamten Arbeitslosigkeit des Mo-

nats berechnet. Dies ist problemlos möglich, da für das gesamte Jahr 1994 die Arbeitslosigkeit nach Bezirken und Monaten vorliegt. Damit unterstellt man jedem Bezirk eine Einwohnerstruktur, die ähnlichen Beschäftigungen nachgeht und damit denselben saisonalen Schwankungen unterliegt. Diese Annahme erscheint plausibel, da manche Bezirke als typische Arbeiterbezirke gelten, andere dagegen eher von Akademikern bevorzugt werden.

Unterstellt man dagegen, dass jeder Bezirk in sich heterogen und der Anteil an der gesamten Arbeitslosigkeit im Mittel gleich ist, so kann man einen Durchschnitt über alle monatlichen Anteile bilden.

Diese beiden Möglichkeiten gehen davon aus, dass die Arbeitslosigkeit des Jahres 1994 der beste Schätzer für die Arbeitslosigkeit 1993 ist. Möglich wäre auch, den Anteil eines Bezirks an der Arbeitslosigkeit für alle Jahre 1994 bis 2000 zu berechnen und einen Durchschnitt über die Jahre zu bilden. Geht man aber wiederum von typischen Arbeiter- oder Akademikerbezirken aus, so unterliegen diese nicht nur denselben saisonalen, sondern auch denselben konjunkturellen Einflüssen. Durch die Durchschnittsbildung über mehrere Jahre würden diese Informationen verloren gehen.

Der nach dem zweiten Schritt berechnete Anteil der Arbeitslosigkeit eines Bezirks wird mit dem Jahresdurchschnitt des Jahres 1993 multipliziert, sofern über die Monate ein Durchschnitt gebildet wurde, bzw. mit den monatlichen Arbeitslosenzahlen, sofern die Anteile pro Monat und Bezirk berechnet wurden.

Zu beachten ist, dass dieser zweite Schritt getrennt für Ost- und West-Berlin erfolgt. Dies hängt damit zusammen, dass die Arbeitslosigkeit in Ost-Berlin nach der Wende höher war als die Bevölkerungszahl implizieren würde. Der Anteil der Arbeitslosen in Ost-Berlin an der Gesamt-Berliner Arbeitslosigkeit fiel von 48,37 % im Jahre 1991 auf 35,41 % im Jahre 1995. Im Mittel betrug er im Zeitraum 1994 bis 2000: 37 % und entsprach damit dem Anteil der Bevölkerung Ost-Berlins an der Bevölkerung Gesamt-Berlins.

Die so für 1993 geschätzten Arbeitslosenzahlen werden mit den tatsächlichen Arbeitslosenzahlen des Jahres 1993 verglichen, soweit vorhanden. Um die Güte der Schätzung zu beurteilen, wurde ein Abweichungsindex der Form

$$\text{Index} = \frac{\text{tatsächliche Anzahl an Arbeitslosen}}{\text{geschätzte Anzahl an Arbeitslosen}}$$

und der Mean Square Error (MSE) aus

MSE =

$$\frac{1}{m}(\text{tatsächliche Anzahl Arbeitslose} - \text{geschätzte Anzahl Arbeitslose})^2$$

berechnet, wobei m = Anzahl der Beobachtungen. Für die Ergebnisse s. Tabelle 7.6.

Tabelle 7.6: *Abweichungsindex und MSE*

Obs	Mean	Std. Dev.	Min	Max	MSE
1. Durchschnittliche Abweichung und monatliche Anteile					
197	1,000180	0,0436	0,908452	1,093130	166.958,8
2. Durchschnittliche Abweichung und mittlerer Anteil des Jahres 1994					
197	1,000046	0,0417	0,900043	1,085894	149.488,7
3. Durchschnittl. Abweichung und mittlerer Anteil über die Jahre 1994 bis 2000					
276	0,975250	0,0867	0,847362	1,127320	581.092,3
4. Abweichung für 1994 und monatliche Anteile					
197	0,996018	0,0434	0,904672	1,088582	163.937,2
5. Abweichung für 1994 und mittlerer Anteil des Jahres 1994					
197	0,995885	0,0415	0,896298	1,081375	146.093,5
6. Abweichung für 1994 und mittlerer Anteil über die Jahre 1994 bis 2000					
276	0,971192	0,0863	0,843836	1,122629	589.183,6

Dabei zeigt sich, dass das beste Ergebnis erzielt wird, wenn die monatlichen Anteile des Jahres 1994 gemittelt werden (Möglichkeiten 2 und 5), und das schlechteste bei der Durchschnittsbildung über die Jahre 1994 bis 2000 (Möglichkeiten 3 und 6). Dies deutet darauf hin, dass sich der Anteil der Arbeitslosen eines Bezirks an allen Berliner Arbeitslosen über die Zeit ändert. Möglich wäre, dass die oben erwähnte homogene Struktur eines Bezirks zwar nicht dazu führt, dass die Arbeitslosigkeit eines Bezirks denselben saisonalen Einflüssen unterliegt, aber denselben konjunkturellen Schwankungen.

Die Bereinigung mit der durchschnittlichen Abweichung der Zahlen des Landesarbeitsamts von den Zahlen der einzelnen Arbeitsämter führt dazu, dass die Arbeitslosigkeit etwas geringer geschätzt wird, d. h. der Abweichungsindex ist größer als 1, als wenn man mit der Abweichung des Jahres 1994 bereinigt. Obwohl die Möglichkeit "Bereinigung mit Abweichung für 1994 und mittlerer Anteil des Jahres 1994" (Möglichkeit 5) im Durchschnitt schlechter ausfällt als die anderen Möglichkeiten (außer 3 und 6), werden doch diese Zahlen weiterhin benutzt und die anderen Möglichkeiten verworfen, da hier der MSE am geringsten ist.

Tabelle 7.7: *Anzahl Arbeitslose im Jahresdurchschnitt 1991 - 2000*

Jahr	Arbeitslose
1991	181.002
1992	205.029
1993	203.211
1994	208.579
1995	212.347
1996	235.323
1997	264.971
1998	272.539
1999	267.483
2000	264.526
<i>Anmerkung:</i> Zahlen bis einschließlich 1993 sind geschätzt. Zahlen ab 1994 von den Berliner Arbeitsämtern.	

Die vollständigen Arbeitslosenzahlen finden sich in Tabelle 7.7. Dabei sind die fehlenden Zahlen bis einschließlich 1993 nach dem eben geschilderten Verfahren geschätzt worden. Soweit tatsächliche Werte vorhanden waren, wurden diese anstelle der geschätzten Zahlen benutzt. Bei den Zahlen ab 1994 handelt es sich um die tatsächlichen Zahlen der Arbeitsämter Berlins, die sich auch in Tabelle 7.5 auf Seite 56 finden.

Neukölln hatte im Durchschnitt die meisten Arbeitslosen, nämlich 24.994. Mit durchschnittlich 2.884 Arbeitslosen lebten die wenigsten in Zehlendorf. Bezogen auf die Bevölkerungsgröße hatte mit durchschnittlich 11,13 % Kreuzberg die meisten Arbeitslosen, die wenigsten mit 2,93 % Zehlendorf (s. Tabelle 7.8 auf Seite 61). Durchschnittlich betrug der Anteil der Arbeitslosen an der Bevölkerung in Berlin 6,74 % (s. Tabelle 7.9 auf Seite 62). Der Anteil der Arbeitslosen an der Bevölkerung unterscheidet sich von der offiziellen Arbeitslosenquote, die aus dem Verhältnis der Arbeitslosen zum Arbeitskräftepotential, das sich wiederum aus Arbeitslosen zuzüglich unselbständig beschäftigten Arbeitnehmern zusammensetzt, berechnet wird. Die offizielle Arbeitslosenquote ist immer höher, da das Arbeitskräftepotential immer geringer als die Einwohnerzahl ist. Es wurde ein der offiziellen Arbeitslosenquote angenäherter Anteil berechnet, indem die Anzahl der Arbeitslosen zur Bevölkerung im erwerbsfähigen Alter in Beziehung gesetzt wurde. Dieser Anteil ist ebenfalls in Tabelle 7.8 ausgewiesen. Der Anteil der Arbeitslosen an

der erwerbsfähigen Bevölkerung im Alter von 15 bis unter 65 Jahren betrug durchschnittlich 9,44 %. In Kreuzberg war er mit 15,01 % am höchsten, in Zehlendorf mit 4,34 % am geringsten.

Tabelle 7.8: Mittlere Anzahl Arbeitslose sowie Anteil an Erwerbsbevölkerung und Gesamtbevölkerung pro Bezirk für den Zeitraum 1991 - 2000

Bezirk	Anzahl	Anteil an Erwerbs- bevölkerung	Anteil an Gesamt- bevölkerung
Mitte	5.107	8,79 %	6,55 %
Tiergarten	7.056	10,40 %	7,73 %
Wedding	14.100	12,05 %	8,57 %
Prenzlauer Berg	12.596	11,86 %	8,91 %
Friedrichshain	8.364	10,81 %	7,99 %
Kreuzberg	16.984	15,01 %	11,13 %
Charlottenburg	11.262	8,64 %	6,26 %
Spandau	14.151	9,12 %	6,32 %
Wilmerdorf	6.189	6,11 %	4,34 %
Zehlendorf	2.884	4,34 %	2,93 %
Schöneberg	11.318	10,03 %	7,43 %
Steglitz	8.541	6,43 %	4,46 %
Tempelhof	9.739	7,42 %	5,12 %
Neukölln	24.994	11,37 %	8,05 %
Treptow	7.023	9,18 %	6,42 %
Köpenick	7.419	9,51 %	6,66 %
Lichtenberg	11.013	9,05 %	6,75 %
Weißensee	4.070	9,34 %	6,63 %
Pankow	7.496	9,17 %	6,52 %
Reinickendorf	13.475	7,85 %	5,40 %
Marzahn	11.478	10,12 %	7,52 %
Hohenschönhausen	7.815	9,39 %	6,74 %
Hellersdorf	8.429	9,35 %	6,40 %

Tabelle 7.9 auf Seite 62 weist den Anteil der Arbeitslosen an der Erwerbsbevölkerung und Gesamtbevölkerung für Gesamt-Berlin aus.

Tabelle 7.9: *Anteil Arbeitslose an Erwerbsbevölkerung und Gesamtbevölkerung für Gesamt-Berlin*

Jahr	Arbeitslose
Anteil an Erwerbsbevölkerung	9,44 %
Anteil an Gesamtbevölkerung	6,74 %

Wie man in der Abbildung 7.2 sieht, stieg der Anteil der Arbeitslosen an der Bevölkerung im Zeitablauf von 5,25 % im Jahr 1991 auf 7,82 % im Jahr 2000 (schwarze Kurve). Die Standardabweichung der Bezirke ist ebenfalls eingezeichnet (blaue Kurve). Diese berechnet sich aus der Wurzel aus der mittleren quadratischen Abweichung des Arbeitslosenanteils eines Bezirks vom Anteil der Arbeitslosen an der Bevölkerung in Berlin.

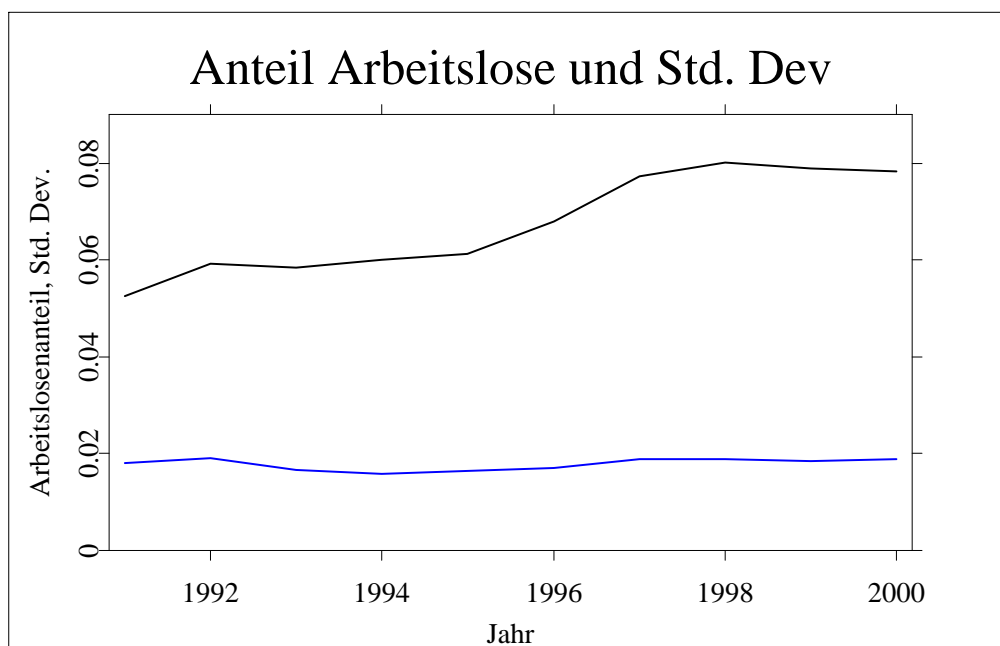


Abbildung 7.2: Mittlerer Anteil Arbeitslose an der Bevölkerung (schwarze Kurve) und Standardabweichung der Bezirke (blaue Kurve)

Kapitel 8

Regression

In diesem Kapitel wird zur eigentlichen Regression übergegangen. Zunächst wird die Vorgehensweise vorgestellt, danach erfolgt die Darstellung der Ergebnisse, erst für die Regression über den gesamten Zeitraum 1991 bis 2000, dann für den Teilzeitraum 1996 bis 2000. Dabei werden jeweils als erstes die Residuen und Ausreißer analysiert, bevor die Koeffizienten besprochen werden. Die verschiedenen Regressionen werden zum Schluss miteinander verglichen.

8.1 Vorgehensweise

Im folgenden werden die Koeffizienten folgender Gleichung geschätzt:

$$\ln(\text{Preis}) = \beta_1 + \sum_{k=1}^K \beta_k x_k + \sum_{i=1}^I \gamma_i z_i + \sum_{d=1}^D \delta_d D_d + \sum_{j=1}^J \tau_j t_j + \epsilon \quad (8.1)$$

wobei

- β_1 - Regressionskonstante
- β_k - Regressionskoeffizienten der Hauseigenschaften, $k = 1, 2, \dots, K$
- x_k - Hausvariablen, $k = 1, 2, \dots, K$
- γ_i - Regressionskoeffizienten der Bezirkseigenschaften, $i = 1, 2, \dots, I$
- z_i - Bezirksvariablen, $i = 1, 2, \dots, I$
- δ_d - Regressionskoeffizienten für Dummy-Variablen, $d = 1, 2, \dots, D$
- D_d - Dummy-Variablen, $d = 1, 2, \dots, D$
- τ_j - Regressionskoeffizienten für Jahresdummies, $j = 1, 2, \dots, J$

t_j - Jahresdummies, $j = 1, 2, \dots, J$
 ϵ - Störgröße

Die abhängige Variable ist der Logarithmus des Preises, da dann die Koeffizienten der Schätzung, wie in Abschnitt 5.2.9 dargestellt, als prozentuale Preisauf- bzw. Preisabschläge interpretiert werden können. Die metrischen Variablen werden zum größten Teil logarithmiert. Da es jedoch keinen Logarithmus von 0 gibt, werden die Variablen **Alter**, **Anteil Waldfläche**, **U-Bahnhöfe/10.000 Einwohner**, **S-Bahnhöfe/100 ha** sowie **Spielfläche/Kind unter 6** nicht logarithmiert.

In Abbildung 8.1 ist der Logarithmus des Preises gegen den Logarithmus der Grundstücksfläche und der Geschossfläche sowie gegen das Alter geplottet.

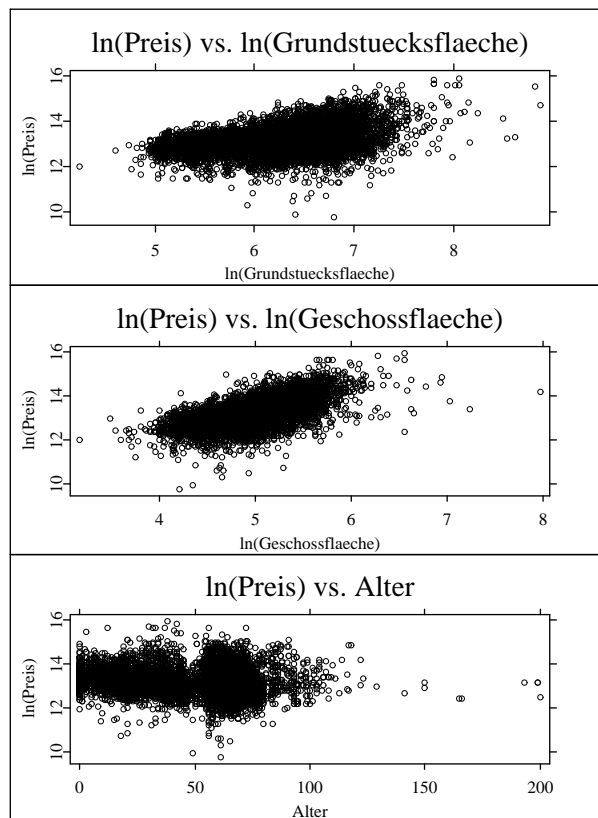


Abbildung 8.1: $\ln(\text{Preis})$ vs. $\ln(\text{Grundstücksfläche})$, $\ln(\text{Preis})$ vs. $\ln(\text{Geschossfläche})$, $\ln(\text{Preis})$ vs. Alter

Man erkennt einen log-log Zusammenhang für die Grundstücks- und die Geschossfläche sowie einen log-linearen Zusammenhang für das Alter.

Insgesamt werden drei metrische Hausvariablen, elf Bezirksvariablen, 43 Dummy-Variablen für die Hauseigenschaften - die sich 16 verschiedenen Gruppen zuordnen lassen - und neun Jahresdummies verwendet. In der Regression über den Teilzeitraum 1996 bis 2000 werden zusätzlich 13 Lagedummies, die sich drei Gruppen zuordnen lassen, benutzt.

Zunächst wurden nur die Hausvariablen betrachtet. Anstelle der Bezirkseigenschaften wurden Bezirksdummies gesetzt, die für die verschiedenen Bezirke kontrollieren. Ziel war es dann, die Bezirksdummies durch konkrete Bezirkseigenschaften zu ersetzen.

Anschließend wurden die Residuen auf Normalverteilung getestet und auf Ausreißer untersucht. Da dies erst nach der Modellschätzung geschehen kann, wurde dafür ein vorläufiges Modell nach der ML-Methode geschätzt. Für das endgültige Modell wurden die Koeffizienten nach der ML-Methode stufenweise und mit robusten Standardfehlern geschätzt.

Für die stufenweise Schätzung müssen zwei Signifikanzniveaus festgelegt werden: eines für die Entfernung aus dem Modell (α_{raus}) und eines für die Hinzufügung zum Modell (α_{rein}), wobei immer $\alpha_{raus} > \alpha_{rein}$ gelten muss. In dieser Arbeit wurde $\alpha_{raus} = 0,10$ und $\alpha_{rein} = 0,05$ gewählt.

Die Schätzung geht so vor sich, dass erst das volle Modell geschätzt wird. Wenn der am wenigsten signifikante Term insignifikant ist, wird er entfernt und das Modell neu geschätzt. Dabei erfolgt ständig ein Abgleich, ob die signifikanteste ausgeschlossene Variable signifikant ist (Variable wird wieder hinzugefügt) und die am wenigsten signifikante eingeschlossene Variable insignifikant ist (Variable wird ausgeschlossen). Das erfolgt so lange, bis alle eingeschlossenen Variablen signifikant sind und alle ausgeschlossenen insignifikant auf dem vorgegebenen Signifikanzniveau.

Diese Vorgehensweise ist für beide Regressionen gleich. Im Abschnitt 8.2 werden die Ergebnisse der Regression über den Gesamtzeitraum 1991 bis 2000 vorgestellt, im Abschnitt 8.3 die Ergebnisse der Regression über den Teilzeitraum 1996 bis 2000.

8.2 Regression auf den Gesamtkaufpreis über den gesamten Zeitraum 1991 bis 2000

8.2.1 Residuen- und Ausreißeranalyse

Die Analyse der Residuen kann, wie bereits erwähnt, erst nach der Modellschätzung vorgenommen werden. Zunächst wurden die Koeffizienten nach der ML-Methode geschätzt. Da es sich dabei um ein vorläufiges Modell handelt, werden die Ergebnisse hier nicht dargestellt.

Danach wurden der Leverage (Hebel), standardisierte Residuen, DFBETA und DFFITS berechnet. Mit diesen Maßzahlen sollen besonders einflussreiche Beobachtungen identifiziert werden.

Geht man vom linearen Regressionsmodell der Form 5.1 aus und schätzt man β nach der ML-Methode, so lautet der Schätzer für y :

$$\hat{y} = X(X'X)^{-1}X'y.$$

Definiert man eine Hatmatrix $H = X(X'X)^{-1}X'$, so lässt sich der Einfluss der i -ten Beobachtung berechnen als

$$h_i = x_i'(X'X)^{-1}x_i.$$

Die h_i sind die Diagonalelemente der Hatmatrix H . Je weiter der Wert eines Regressors von dem Mittelwert der anderen Regressoren entfernt ist, desto größer ist der Leverage:

$$h_i = \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum_{j=1}^n (x_j - \bar{X})^2}. \quad (8.2)$$

Es gilt immer

$$0 \leq h_i \leq 1.$$

Als einflussreich werden diejenigen Beobachtungen angesehen, deren $h_i \geq 2K/n$ ist (vgl. Judge et al. (1998)). Im vorliegenden Datensatz haben 269 Beobachtungen diesen Grenzwert überschritten.

Standardisiert man die Residuen anhand ihrer geschätzten Standardabweichung, erhält man die studentisierten Residuen

$$\hat{\epsilon}_i' = \frac{\hat{\epsilon}_i}{\hat{\sigma}\sqrt{1-h_i}}. \quad (8.3)$$

Die Fälle, bei denen $\hat{\epsilon}_i' \geq 2$, werden als Ausreißer gesehen. Im vorliegenden Datensatz waren das 209 Fälle.

DFFITs versucht, diese beiden Maßzahlen zu einer einzigen zusammenzufassen:

$$DFFITs = \hat{\epsilon}_i' \sqrt{\frac{h_i}{1 - h_i}}. \quad (8.4)$$

Diese Maßzahl misst den Einfluss der i-ten Beobachtung auf die Vorhersage des Modells. Als Grenzwert wird üblicherweise $|DFFITs| > 2\sqrt{\frac{K}{n}}$ genommen (vgl. z. B. Belsley et al. (1980)). 311 Beobachtungen überschritten diesen Grenzwert.

DFBETA gibt den Beitrag der i-ten Beobachtung zum Schätzer $\hat{\beta}$ an und wird aus der Differenz $\hat{\beta} - \hat{\beta}(i)$ berechnet, d. h. dem Wert von $\hat{\beta}$ einschließlich der i-ten Beobachtung abzüglich des Wertes, den $\hat{\beta}$ bei Ausschluss der i-ten Beobachtung annimmt.

$$DFBETA = \hat{\beta} - \hat{\beta}(i) = \frac{(X'X)^{-1}x_i\hat{\epsilon}_i}{1 - h_i}. \quad (8.5)$$

Der Grenzwert ist $|DFBETA_i| > \frac{2}{\sqrt{n}}$ (vgl. z. B. Belsley et al. (1980)). Aufgrund der großen Anzahl unabhängiger Variablen wurden die DFBETAs in dieser Arbeit nur für Hausvariablen berechnet, die signifikant und deren Koeffizienten $\hat{\beta}_k \geq 0,10$ waren. Die Variablen und die Anzahl der Beobachtungen, die den Grenzwert überschritten, sind in der Tabelle 8.1 auf Seite 68 dargestellt.

Gemessen an der Gesamtanzahl an Beobachtungen, die in die Regression eingehen, nämlich 10.136, ist die ausgewiesene Anzahl an Ausreißern gering. Sie werden daher in der Regression belassen.

Die Annahme der Normalverteilung der Residuen ist Voraussetzung für die Anwendung der Maximum-Likelihood-Methode (s. Abschnitt 5.2.6). Die Residuen müssen daher auf Normalverteilung getestet werden. Dafür kann man sich des dritten und vierten Moments einer Verteilung bedienen. Im allgemeinen linearen Modell der Form $y = X\beta + \epsilon$ ist $E[\epsilon] = 0$ und $E[\epsilon'\epsilon] = \sigma^2 I$. Wenn ϵ normalverteilt ist, dann sind das dritte und vierte Moment $\mu_3 = E[\epsilon_n^3] = 0$ und $\mu_4 = E[\epsilon_n^4] = 3\sigma^4$. Werden diese beiden Momente standardisiert, erhält man die Schiefe (Skewness) bzw. die Wölbung (Kurtosis).

Tabelle 8.1: Variablen x_k und Anzahl Beobachtungen, für die $DFBETA_{ki} > \frac{2}{\sqrt{n}}$

Variable	Anzahl
Grundstücksfläche	326
Geschossfläche	359
Geschäftliche Bindung	38
Persönliche Verhältnisse	85
Wasserlage	37
Berliner Dach	9
Walmdach	356
Erwerbergruppe: sonst. jur. Person	94
Zustandsnote: gut	281
Zustandsnote: schlecht	302
Erwerbergruppe: sonstige	5
Veräußerergruppe: GbR	23
Veräußerergruppe: BRD	77
Veräußerergruppe: Landeseigene Wohnungsunternehmen	22
Mieterkauf	181
Vermietet	74

Schiefe Verteilungen können linksschief/rechtssteil oder rechtsschief/linkssteil sein. Bei einer symmetrischen Verteilung ist die Skewness = 0. Sie wird durch

$$\sqrt{b_1} = \frac{\mu_3}{\sigma^4}$$

gemessen.

Die Kurtosis misst die Wölbung der Kurve:

$$b_2 = \frac{\mu_4}{\sigma^4}$$

(vgl. Judge et al. (1998))

Der Test auf Normalverteilung erfolgte hier mit Hilfe des von D'Agostino et al. (1990) vorgeschlagenen Schätzers. Die Nullhypothese auf Normalverteilung musste abgelehnt werden.

Abbildung 8.2 auf Seite 69 enthält den Residualplot. Da die Residuen nicht normalverteilt sind, werden die Standardfehler robust geschätzt. Die Ergebnisse der Schätzung sind in den folgenden Abschnitten dargestellt.

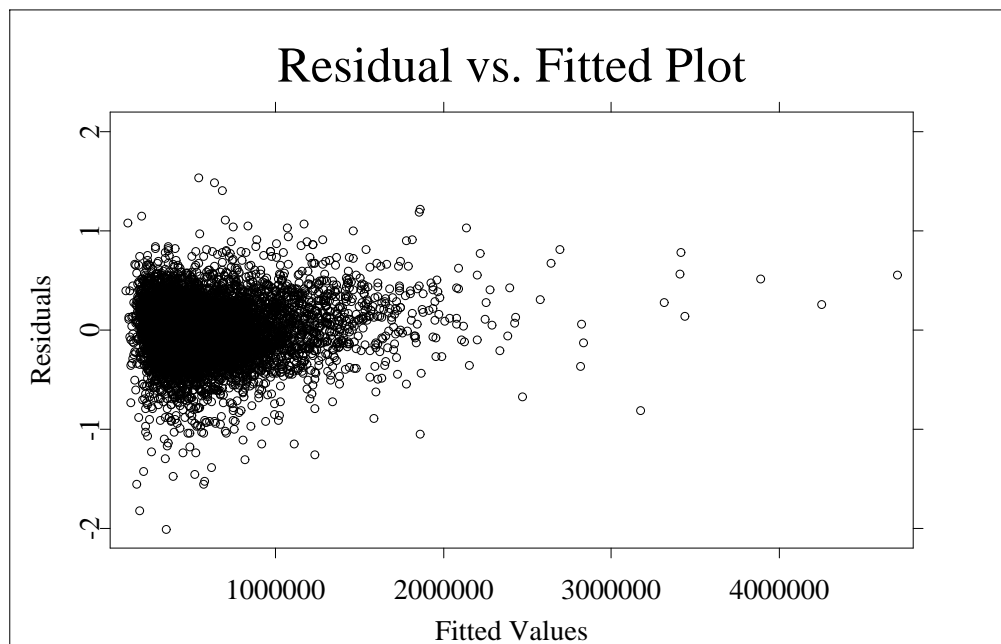


Abbildung 8.2: Residual vs. Fitted Plot für den Gesamtzeitraum 1991 - 2000

8.2.2 Bezirksdummies

Wie in Abschnitt 8.1 bereits erwähnt, wurden zunächst in der Gleichung 8.1 anstelle des Terms $\sum_{i=1}^I \gamma_i z_i$ Bezirksdummies gesetzt, die für das allgemeine Preisniveau in den Bezirken kontrollieren sollen. Erst im zweiten Durchgang wurden diese durch konkrete Bezirkseigenschaften ersetzt.

Die Regressionskennzahlen können Tabelle 8.2 entnommen werden. Da die Standardfehler robust geschätzt werden, kann die Analysis of Variance nicht ausgewiesen werden.

Tabelle 8.2: *Regressionskennzahlen: Modell mit Bezirksdummies*

Number of Obs	10.136
F(54, 10.081)	429,39
Prob>F	0,0000
R^2	0,7434
\bar{R}^2	0,7425
Root MSE	0,2602

In Tabelle 8.3 sind die Koeffizienten der Bezirksdummies aufgelistet, um das relative Preisniveau der Bezirke untereinander darzustellen. Da es sich um eine vorläufige Regression handelt, wurde auf die Darstellung der Koeffizienten der anderen Variablen verzichtet. Der Referenzbezirk ist Tempelhof, weil hier im Betrachtungszeitraum die meisten Häuser verkauft wurden. In Mitte, Tiergarten, Friedrichshain, Kreuzberg und Schöneberg wurde im Betrachtungszeitraum kein Haus verkauft. Daher konnten für diese Bezirke keine Koeffizienten geschätzt werden. Der Koeffizient von Reinickendorf war insignifikant.

Tabelle 8.3: *Koeffizienten der Bezirksdummies*

Variable	Coef.	t	$P > t $
Wedding	-0,294	-2,60	0,009
Prenzlauer Berg	-0,164	-2,00	0,046
Charlottenburg	0,350	13,13	0,000
Spandau	0,123	-13,77	0,000
Wilmerdorf	0,637	13,74	0,000
Zehlendorf	0,424	41,10	0,000
Steglitz	0,123	12,16	0,000
Neukölln	-0,064	-7,76	0,000
Treptow	-0,451	-18,58	0,000
Köpenick	-0,239	-14,92	0,000
Lichtenberg	-0,329	-10,45	0,000
Weißensee	-0,433	-19,53	0,000
Pankow	-0,329	-22,50	0,000
Marzahn	-0,425	-28,68	0,000
Hohenschönhausen	-0,392	-20,67	0,000
Hellersdorf	-0,339	-19,38	0,000

Anmerkung: Die abhängige Variable ist der Logarithmus des Preises. Die Schätzung erfolgte gem. Gleichung 8.1 stufenweise und robust nach der Maximum-Likelihood-Methode.

Die Ergebnisse der Tabelle 8.3 stimmen mit denen der Tabelle 6.2 überein: Der Preisaufschlag ist für Wilmerdorf mit 63,7 % am höchsten, gefolgt von Zehlendorf und Charlottenburg. Der größte Preisabschlag erfolgt für Treptow mit 45,1 %. Danach folgen Weißensee und Marzahn. Der Unterschied ist, dass hier bereits für die verschiedenen Hauseigenschaften und Jahre kontrolliert wurde.

In den folgenden Abschnitten werden die Koeffizienten der gesamten Regression nach Gleichung 8.1 vorgestellt. Die Bezirksdummies werden durch

Bezirkseigenschaften ersetzt.

8.2.3 Regressionskennzahlen

Wie man Tabelle 8.4 entnehmen kann, werden immerhin 73,74 % der Variation von y durch das Modell erklärt. Von den insgesamt 11.503 Beobachtungen wurden 10.136 für die Regression benutzt. Die Differenz von 1.367 Beobachtungen fiel aufgrund von missings aus der Regression heraus.

Tabelle 8.4: *Regressionskennzahlen: Modell mit Bezirksvariablen*

Number of Obs	10.136
F(49, 10.086)	437,78
Prob>F	0,0000
R^2	0,7374
\overline{R}^2	0,7369
Root MSE	0,2632

Das in Tabelle 8.4 ausgewiesene Bestimmtheitsmaß ist etwas geringer als das in Tabelle 8.2 ausgewiesene. Damit wurden die Bezirksdummies fast vollständig durch die gefundenen Bezirksvariablen erklärt, ohne dass das Modell jedoch an Erklärungskraft gewonnen hätte.

8.2.4 Koeffizienten der Hauseigenschaften und deren Interpretation

Die Koeffizienten der Hauseigenschaften können Tabelle 8.5 entnommen werden. Bei der folgenden Regression ist ein bezugsfreies Einzelhaus auf einem Frontgrundstück mit einem Vollgeschoss und Satteldach in normalem Zustand, von privat an privat im Jahre 1991 verkauft, das Referenzhaus. Als Referenzeigenschaften wurden die Ausprägungen gewählt, die am häufigsten vorkamen (vgl. Tabelle 6.4 auf Seite 43).

Tabelle 8.5: *Koeffizienten der Hauseigenschaften*

Variable	Coef.	t	P> t
Alter	-0,003	-17,58	0,000
ln(Grundstücksfläche)	0,295	32,18	0,000
ln(Geschossfläche)	0,498	34,94	0,000
Dachausbau	0,034	4,53	0,000
Geschäftliche Bindung	0,270	2,29	0,022

Fortsetzung auf der nächsten Seite

Tabelle 8.5: *Koeffizienten der Hauseigenschaften - Fortsetzung*

Variable	Coef.	t	P> t
Persönl. Verhältnisse des Geschäftsverkehrs	-0,206	-8,34	0,000
Ungewöhnl. Verhältnisse am Grundstück	-0,090	-4,57	0,000
Ungewöhnl. Verhältnisse des Geschäftsverkehrs	-0,082	-2,79	0,005
Wasserlage	0,325	7,67	0,000
Hammergrundstück	-0,040	-3,01	0,003
Pultdach	-0,039	-3,02	0,003
Walmdach	0,106	12,84	0,000
Mansardendach	0,037	2,05	0,041
Zeltdach	0,078	4,18	0,000
Erwerber: GbR	0,063	1,93	0,054
Erwerber: sonst. jur. Person	0,170	6,14	0,000
Zustand: gut	0,139	17,96	0,000
Zustand: schlecht	-0,207	-16,55	0,000
Doppelhaushälfte	-0,062	-9,46	0,000
Reihenendhaus	-0,069	-7,05	0,000
Zwei Vollgeschosse	0,060	6,93	0,000
Drei oder vier Vollgeschosse	0,051	1,81	0,070
Veräußerer: GbR	0,137	3,25	0,001
Veräußerer: BRD	-0,149	-4,71	0,000
Veräußerer: Gemeinn. Wohnunt.	-0,091	-6,31	0,000
Veräußerer: sonst. jur. Person	0,064	5,34	0,000
Veräußerer: Landeseigene Wohnungsunternehmen	-0,179	-4,95	0,000
Mieterkauf	-0,145	-7,49	0,000
Vermietet	-0,231	-7,52	0,000
<i>Anmerkung:</i> Die abhängige Variable ist der Logarithmus des Preises. Die Schätzung erfolgte gem. Gleichung 8.1 stufenweise und robust nach der Maximum-Likelihood-Methode. Die Variablen sind im Anhang A.1 erklärt.			

Das Alter hat erwartungsgemäß ein negatives Vorzeichen. Steigt das Alter des Hauses um ein Jahr, so sinkt der Preis um 0,3 %. Nun ist aber nicht nur interessant, wie sich der Preis ändert, wenn das Alter absolut um ein Jahr steigt, sondern auch, wie sich der Preis bei einer relativen Änderung des Alters um beispielsweise 1 % verhält. So fällt ein zusätzliches Jahr bei einem Neubau stärker ins Gewicht als bei einem Haus, das bereits hundert Jahre alt ist. In den ersten Jahren wird der Preis daher stärker fallen. Dafür muss die Elastizität berechnet werden. Setzt man in Gleichung 5.12 auf Seite 37 für x_k das mittlere Alter von 40,57 Jahren (vgl. Tabelle 6.1 auf Seite 40) ein, so ergibt sich eine Preisänderung von 0,12 %, wenn das durchschnittliche Haus

um 0,4057 Jahre oder 4,8684 Monate altert.

Für die Grundstücksfläche und die Geschossfläche wurden die logarithmierten Werte betrachtet. Steigt die Grundstücksfläche um 1 %, wird ein Preisaufschlag von 0,295 % gezahlt. Analog betrug der Preisaufschlag für 1 % mehr Geschossfläche 0,498 %.

Hat das Haus anstelle nur eines Vollgeschosses zwei, so beträgt der Preisaufschlag 6,0 %. Bei einem dritten oder vierten Vollgeschoss beträgt er 5,1 % gegenüber einem Haus mit nur einem Vollgeschoss. Obwohl ein Haus mit nur einem Vollgeschoss am häufigsten anzutreffen ist (s. Tabelle 6.4), scheinen Häuser mit zwei Vollgeschossen am meisten geschätzt zu werden. Dagegen wird für ein ausgebautes Dachgeschoss nur ein Aufschlag von 3,4 % gezahlt. Ein zusätzliches Vollgeschoss wird gegenüber einem ausgebauten Dachgeschoss mehr geschätzt. Dies ist plausibel, da ein Dachgeschoss wegen seiner Schrägen nicht in dem Umfang genutzt werden kann wie ein zusätzliches Vollgeschoss.

Liegt das Haus am Wasser, beträgt der Preisaufschlag sogar 32,5 %. Dies deckt sich mit den Ergebnissen von Luttik (2000), der einen Preisaufschlag von bis zu 28 % für ein Haus mit Wasserzugang feststellte.

Sobald ungewöhnliche Verhältnisse vorliegen, sinkt der Preis. Bei persönlichen Verhältnissen des Geschäftsverkehrs beträgt der Abschlag 20,6 %, bei ungewöhnlichen Verhältnissen tatsächlicher Art 9,0 % und bei ungewöhnlichen Verhältnissen des Geschäftsverkehrs 8,2 %. Bei geschäftlicher Bindung wird dagegen ein Preisaufschlag von 27,0 % gezahlt.

Von den verschiedenen möglichen Blocklagen gibt es nur für das Hammergrundstück einen Preisabschlag von 4,0 %. Die Auf- bzw. Abschläge für ein Hinterland-, Mehrfronten- oder Eckgrundstück sind dagegen insignifikant gegenüber dem üblichen Frontgrundstück.

Hat das Haus ein Pultdach anstelle eines Satteldaches, dann sinkt der Preis um 3,9 %. Für ein Walm-, Mansarden- oder Zeltdach wird dagegen ein Aufschlag von 10,6 %, 3,7 % bzw. 7,8 % gezahlt. Die anderen Dachformen sind insignifikant.

Gesellschaften bürgerlichen Rechts und sonstige juristische Personen zahlen 6,3 % bzw. 17 % mehr für Häuser als Privatpersonen. Tritt die Bundesrepublik Deutschland, ein (gemeinnütziges) Wohnungsunternehmen oder ein

Landeseigenes Wohnungsunternehmen als Veräußerer auf, so wird ein Preisabschlag von 14,9 %, 9,1 % bzw. 17,9 % gezahlt. Einen Preisaufschlag verlangen nur die Gesellschaften bürgerlichen Rechts (in Höhe von 13,7 %) und sonstige juristische Personen (in Höhe von 6,4 %).

Erwartungsgemäß wird für ein Haus in gutem Zustand ein Aufschlag von 13,9 % gezahlt, während der Abschlag für ein Haus in schlechtem Zustand 20,7 % gegenüber einem Haus in normalem Zustand beträgt.

Für eine Doppelhaushälfte beträgt der Preisabschlag 6,2 %, für ein Reihendhaus 6,9 %. Trotzdem ist der Preisunterschied eines Reihenhauses insignifikant gegenüber einem einzelnstehenden Haus.

Handelt es sich um einen Mieterkauf, so beträgt der Abschlag 14,5 %, ist das Haus vermietet, beträgt er 23,1 %. Dies könnte ein Beleg dafür sein, dass Einfamilienhäuser überwiegend selbst genutzt werden und nicht als Renditeobjekt dienen. Bei vermieteten Häusern stellt sich die Frage, innerhalb welches Zeithorizonts der Mieter auszieht. Der Abschlag wäre dann eine Entschädigung für den entgangenen Nutzen während der Kündigungsfrist.

8.2.5 Koeffizienten der Bezirkseigenschaften und deren Interpretation

Soweit es sich um Hauseigenschaften handelt, entsprechen die Vorzeichen der Koeffizienten also weitestgehend den Erwartungen. Unerwartete Vorzeichen gibt es bei den Bezirkseigenschaften. Deren Koeffizienten können Tabelle 8.6 entnommen werden.

Tabelle 8.6: *Koeffizienten der Bezirkseigenschaften*

Variable	Coef.	t	P> t
Anteil Waldfläche	0,141	3,92	0,000
ln(Baum)	0,067	3,37	0,001
Anzahl U-Bahnhöfe pro 10.000 Einwohner	0,307	13,73	0,000
Anzahl S-Bahnhöfe pro 100 Hektar	0,167	2,41	0,016
ln(Bevölkerung unter 6/Grundschulklasse)	0,202	8,25	0,000
ln(Spielfläche/Kind unter 15)	-0,034	-2,56	0,010
ln(Anteil Kinder)	-0,166	-3,34	0,001
ln(Nichterwerbsperson/Erwerbsperson)	-0,128	-2,14	0,032
ln(Arbeitslose/Erwerbsperson)	-0,384	-10,37	0,000
ln(Haushaltsnettoeinkommen)	0,347	5,08	0,000

Fortsetzung auf der nächsten Seite

Tabelle 8.6: *Koeffizienten der Bezirkseigenschaften - Fortsetzung*

Variable	Coef.	t	P> t
ln(Ausländer/Einwohner)	0,111	7,56	0,000
<i>Anmerkung:</i> Die abhängige Variable ist der Logarithmus des Preises. Die Schätzung erfolgte gem. Gleichung 8.1 stufenweise und robust nach der Maximum-Likelihood-Methode. Die Variablen sind im Anhang A.1 erklärt.			

In Abschnitt 4.3 wurde dargelegt, von welchen Eigenschaften eines Bezirks man annimmt, dass sie die Kaufentscheidung und den Kaufpreis eines bebauten Grundstücks beeinflussen. Die Variablen können allerdings nicht einfach alleinstehend übernommen werden. So ist es beispielsweise sinnvoller, anstelle der absoluten Waldfläche eines Bezirks den Anteil der Waldfläche an der Bezirksfläche zu nehmen, da ein großer Bezirk mehr Waldfläche haben kann als ein kleiner Bezirk. Im folgenden soll der Einfluss von Wald und Bäumen, der Infrastruktur, der sozialen Infrastruktur, der sozio-ökonomischen Aspekte, der Altersstruktur und der Ausländer untersucht werden.

Als ein Aspekt beim Kauf eines Einfamilienhauses wurde in dieser Arbeit die Verwirklichung des Traumes von einem "Haus im Grünen" gesehen. Die entsprechende Hypothese lautete, dass der Preis eines bebauten Grundstücks umso höher ist, je "grüner" ein Bezirk ist. Es zeigte sich, dass die Variablen **Anteil Waldfläche** und **ln(Baum)** dies am besten erfassen. **Anteil Waldfläche** gibt an, welcher Anteil der Bezirksfläche mit Bäumen oder Sträuchern bedeckt ist. Diese Variable wird nicht logarithmiert, da es sieben Bezirke ohne Waldfläche gibt. Erwartungsgemäß ist der Preis in Bezirken mit viel Waldfläche höher, so dass der Koeffizient positiv ist. Die durchschnittliche Waldfläche in Berlin betrug 17,88 %. Steigt der Anteil der Waldfläche, ausgehend von diesem Wert, um 1 %, so steigt der Preis um 0,025 %. Ebenso steigt der Preis um 0,067 %, wenn die Anzahl der Bäume pro 1 km Straßenlänge um 1 % zunimmt.

Für die Infrastruktur wurde angenommen, dass der Preis in Bezirken mit guter Verkehrsanbindung höher sein sollte. Die Anzahl der U-Bahnstationen war dann signifikant, wenn sie zur Anzahl der Einwohner eines Bezirks in Beziehung gesetzt wurde. Die Anzahl der S-Bahnstationen war in Beziehung zur Bezirksfläche signifikant. Dies könnte daran liegen, dass das U-Bahn-Netz in den Innenstadtbezirken besonders dicht ist, während die S-Bahn gerade die Randbezirke mit dem Zentrum verbindet und damit eher ein Flächenverkehrsmittel ist (s. Abbildung 6.3 auf Seite 48). Es wurde wieder auf die Logarithmierung verzichtet, da nicht alle Bezirke U- oder S-Bahnanschluss

aufweisen. Wie erwartet, sind die Koeffizienten positiv. Steigt die Anzahl der U-Bahnhöfe pro 10.000 Einwohner um 1, so steigt der Preis um 30,7 %. Analog steigt er bei einem zusätzlichen S-Bahnhof pro 100 Hektar um 16,7 %.

Die Anzahl der Grundschulklassen wurde zur Bevölkerung unter sechs Jahren in Beziehung gesetzt, indem die Anzahl der Kinder unter sechs Jahren pro Grundschulklasse berechnet wurde. Damit wird unterstellt, dass die Anzahl der in den nächsten Jahren einzuschulenden Kinder pro Grundschulklasse für die Kaufentscheidung wichtiger ist als die jetzige Anzahl der Kinder pro Grundschulklasse. Die zugrundeliegende Annahme ist, dass die Anpassung der Grundschulklassen an die Größe der Schülergeneration zeitverzögert erfolgt. Die entsprechende Hypothese lautete: je geringer die zukünftige Klassengröße, desto höher der Preis. Wider Erwarten ist der entsprechende Koeffizient positiv, d. h. steigt die Anzahl der Kinder unter sechs Jahren pro Grundschulklasse um 1 %, so steigt der Preis eines bebauten Grundstücks um 0,202 %. Ähnliche Ergebnisse werden auch dann erzielt, wenn anstelle der Kinder unter sechs Jahren die Altersgruppe der Kinder von sechs bis unter 15 Jahren betrachtet wird. Der Quotient ist dann hoch, wenn es bei gegebener Anzahl Grundschulen viele Kinder unter sechs Jahren gibt oder umgekehrt bei gegebener Anzahl an Kindern wenige Grundschulen. Wie man am Koeffizienten für $\ln(\text{Anteil Kinder})$ sieht, hat der Kinderanteil einen negativen Einfluss auf die Preise. Die positive Wirkung des Koeffizienten muss dann von der geringen Anzahl von Grundschulen herrühren. Vielleicht wird diese als Signal dafür genommen, wie viele Kinder es im Bezirk gibt.

Für die Spielfläche wurde angenommen, dass die Preise sinken, je mehr zur Verfügung steht, da sie vor allem in Bezirken mit dichter Bebauung eingerichtet werden. Der Koeffizient der Variable $\ln(\text{Spielfläche pro Kind unter 15})$ ist den Erwartungen entsprechend negativ: der Preis sinkt um 0,034 %, wenn die Spielfläche pro Kind unter 15 Jahren um 1 % steigt. Damit ist die genannte Hypothese bestätigt.

Der sozio-ökonomische Aspekt wird anhand der Variablen *Nichterwerbspersonen pro Erwerbsperson*, *Anteil der Arbeitslosen an der Erwerbsbevölkerung* und *Mittleres monatliches Haushaltsnettoeinkommen* gemessen. Der Quotient *Nichterwerbsperson/Erwerbsperson* gibt an, wie viele Nichterwerbspersonen durch eine Erwerbsperson unterhalten werden und wird, wie im Anhang A Tabelle A.3 erläutert, durch

$$\frac{\text{Bevölkerung unter 15} + \text{Bevölkerung älter als 65}}{\text{Bevölkerung im Alter von 15 bis unter 65 Jahren}}$$

angenähert, da keine Zahlen zu Erwerbspersonen auf Bezirksebene vorliegen. Im Berliner Durchschnitt kommen 0,3985 Nichterwerbspersonen auf eine Erwerbsperson. Die Nichterwerbspersonen setzen sich zu etwa gleichen Teilen aus Kindern und Senioren zusammen (s. Abschnitt 6.2). Steigt der Quotient um 1 %, sinkt der Preis um 0,128 %. Analog sinkt der Preis um 0,384 %, wenn der Anteil der Arbeitslosen an den Erwerbspersonen um 1 % steigt. Dagegen steigt der Preis um 0,347 %, wenn das monatliche Haushaltsnettoeinkommen um 1 % steigt. Damit ist die genannte Hypothese bestätigt, dass ein wohlhabendes Wohnumfeld zu steigenden Preisen führt. Dies könnte an positiven externen Effekten liegen, die mit diesem in Zusammenhang gebracht werden.

Die Anzahl an Ausländern beeinflusst die Kaufentscheidung positiv. Steigt der Ausländeranteil um 1 %, so steigt der Kaufpreis um 0,111 %. Dies widerspricht der genannten Hypothese. Ein Grund könnte die historische Entwicklung sein. So lebten 1991 nur etwa 9,54 % aller Ausländer im Ostteil der Stadt. Das entspricht einem Ausländeranteil von 2,42 % an der Bevölkerung, während er 14,99 % im Westteil betrug. Dieser Anteil stieg auf 19,72 % im Jahr 2000, was einem Anteil von 5,65 % an der Bevölkerung im Ostteil gegenüber einem Anteil von 17,20 % im Westteil entspricht. Nicht zu vergessen ist, dass sich auch das Preisniveau in den beiden Stadthälften nur langsam annähert (vgl. Abschnitt 6.1). Wenn auch die Angleichung der Ausländeranteile wahrscheinlich nicht der Grund für die Angleichung der Preise ist, unterliegen beide Prozesse eventuell einer gemeinsamen Antriebskraft. Zu beachten ist allerdings, dass in den Bezirken mit den höchsten Ausländeranteilen, wie Kreuzberg, Tiergarten und Neukölln, keine Einfamilienhäuser verkauft wurden, und die Beobachtungen in Wedding aufgrund von missings in den meisten Jahren aus der Regression ausgeschlossen wurden. Bei diesen Bezirken handelt es sich gleichzeitig um die Problembezirke Berlins (vgl. Abschnitt 4.1). Der Koeffizient könnte daher verzerrt sein. Hier besteht noch Untersuchungsbedarf. Beispielsweise könnte man den Markt für Eigentums- oder Mietwohnungen untersuchen, da auf diesen Märkten für die Bezirke mit den höchsten Ausländeranteilen eher Beobachtungen vorliegen.

Der Einfluss der Altersstruktur wurde für den Anteil der Kinder an der Gesamtbevölkerung

$$\frac{\text{Bevölkerung unter 6} + \text{Bevölkerung von 6 bis unter 15}}{\text{Gesamtbevölkerung}}$$

überprüft. Der entsprechende Koeffizient ist negativ, was der vorher aufgestellten Hypothese widerspricht. Steigt der Anteil an Kindern um 1 %, so sinkt der Kaufpreis um 0,166 %. Aufgrund dessen, wie der Koeffizient

Nichterwerbsperson/Erwerbsperson gebildet wurde, kontrolliert er auch die Altersstruktur eines Bezirks: Je höher der Anteil der Kinder und der Senioren im Verhältnis zur Bevölkerung zwischen 15 und 65 Jahren ist, desto geringer ist der Preis. Damit ist dieses Ergebnis genau entgegengesetzt zu dem von Walden (1990) gefundenen positiven Einfluss. Untersucht man nur den Einfluss des Seniorenanteils, ist auch dieser Koeffizient negativ. Insgesamt lässt sich ein negativer Einfluss des Kinderanteils auf den Preis feststellen. Dies ist insofern plausibel, als gerade Familien mit Kindern wahrscheinlich nicht das nötige Einkommen haben, um sich ein Haus in einer bevorzugten Wohngegend zu leisten, so dass eher Personen ohne Kinder die Nachfrager stellen und auf diese ein hoher Kinderanteil negativ wirken könnte.

8.2.6 Koeffizienten der Jahresdummies und deren Interpretation

Die Koeffizienten der Jahresdummies können Tabelle 8.7 entnommen werden. Das Jahr 1991 ist das Referenzjahr. Alle anderen Koeffizienten müssen als Zuschläge bezüglich dieses Referenzjahrs interpretiert werden.

Tabelle 8.7: *Koeffizienten der Jahresdummies*

Variable	Coef.	t	P> t
1992	0,197	11,71	0,000
1993	0,180	10,42	0,000
1994	0,222	11,19	0,000
1995	0,221	10,72	0,000
1996	0,188	7,76	0,000
1997	0,194	7,00	0,000
1998	0,195	6,81	0,000
1999	0,159	5,41	0,000
2000	0,106	3,45	0,001
Konstante	3,852	7,44	0,000

Anmerkung: Die abhängige Variable ist der Logarithmus des Preises. Die Schätzung erfolgte gem. Gleichung 8.1 stufenweise und robust nach der Maximum-Likelihood-Methode. Die Variablen sind im Anhang A.1 erklärt.

Die Jahresdummies fangen die allgemeine Preisentwicklung auf, die sich nicht durch die Veränderungen in den Bezirksvariablen erklären lässt. Hierzu gehören z. B. die allgemeine wirtschaftliche Lage, der Zinssatz und Nachfrage und Angebot an Immobilien.

Wie man auch in Abbildung 8.3 sieht, steigen die Preise bedingt durch den

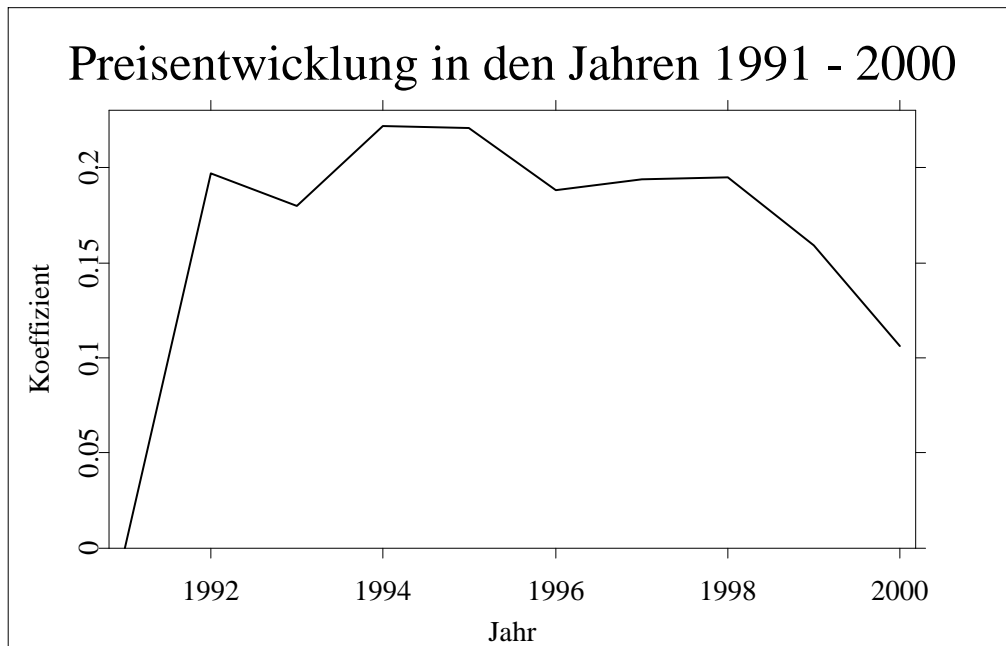


Abbildung 8.3: Preisentwicklung in den Jahren 1991 bis 2000

Immobilienboom durch die Wiedervereinigung zunächst an und fallen ab 1995, wobei sie sich noch bis 1998 ungefähr halten können, um ab 1998 endgültig zu sinken, ohne jedoch das Ausgangsniveau von 1991 wieder zu erreichen.

8.3 Regression auf den Gesamtkaufpreis über den Zeitraum 1996 bis 2000

8.3.1 Vorgehensweise und Ausreißeranalyse

Wie in Abschnitt 7.1 dargestellt, wiesen einige Variablen am Anfang des Betrachtungszeitraums verstärkt missings auf. Ab 1996 wurden folgende Hausvariablen - zusätzlich zu den im Abschnitt 8.2 benutzten - weitgehend vollständig erfasst: **Typische Geschossflächenzahl**, **Typische Nutzungsart** und **Wohnlage**. In diesem Abschnitt soll daher eine Regression über den Teilzeitraum 1996 bis 2000 erfolgen. Die Beobachtungen, die in den Zeitraum 1991 bis 1995 fallen, wurden entfernt. Ebenso wurde eine Beobachtung mit der Dachform "Berliner Dach" entfernt, da eine Beobachtung zu wenig ist, um für diese Ausprägung zu kontrollieren. Der Datensatz umfasste jetzt 6.644 Be-

obachtungen.

Wie bereits in Abschnitt 3.1 dargelegt, beschreiben die Variablen **Typische Nutzungsart**, **Wohnlage** und **Typische Geschossflächenzahl** die unmittelbare Umgebung des Objekts und sind damit individueller als die Bezirksvariablen. Deren Koeffizienten werden daher gesondert dargestellt. Die Anzahl der missings ist in Tabelle 8.8 aufgeführt.

Tabelle 8.8: *Anzahl missings für Lagevariablen*

Variable	Anzahl
Typische Nutzungsart	81
Wohnlage	145
Typische GFZ	152

Die **Typische Geschossflächenzahl** ist eine diskrete Variable mit 14 verschiedenen Ausprägungen zwischen 0 und 4, wobei die meisten Beobachtungen, nämlich 5.235, die Ausprägung 0,4 aufwiesen. Diese Ausprägung wurde daher auch als Referenzpunkt der Regression benutzt. Zusätzlich zu den im Abschnitt 8.1 genannten Referenzpunkten sind jetzt Wohnlage: Mittlere Wohnlage und typische Art der Nutzung: Wohnbebauung Referenzausprägungen. Die Nutzungsarten Dienstleistungsgewerbe (Anzahl der Beobachtungen: 5), produzierendes Gewerbe (3) und mischgebietstypischer Charakter (1) wurden zu einer Ausprägung "sonstige" zusammengefasst.

Wie auch in der vorherigen Regression wurde zunächst eine vorläufige ML-Schätzung vorgenommen und auf Ausreißer überprüft. Einen hohen Leverage wiesen 222 Beobachtungen aus, studentisierte Residuen größer als 2: 117 Beobachtungen und einen erhöhten DFFIT 164 Beobachtungen. Die DFBETAs sind wieder tabellarisch dargestellt (s. Tabelle 8.9).

Tabelle 8.9: *Variablen x_k und Anzahl Beobachtungen, für die $DFBETA_{ki} > \frac{2}{\sqrt{n}}$, für die Teilregression 1996 - 2000*

Variable	Anzahl
Grundstücksfläche	206
Geschossfläche	197
Geschäftliche Bindung	33
Persönliche Verhältnisse	66
Wasserlage	23
Erwerbergruppe: sonst. jur. Person	43

Fortsetzung auf der nächsten Seite

Tabelle 8.9: Variablen x_k und Anzahl Beobachtungen, für die $DFBETA_{ki} > \frac{2}{\sqrt{n}}$ für die Teilregression 1996 - 2000 - Fortsetzung

Variable	Anzahl
Erwerbergruppe: sonstige	4
Zustandsnote: gut	180
Zustandsnote: schlecht	196
Veräußerergruppe: BRD	63
Veräußerergruppe: Gemeinnützige Wohnungsunternehmen	99
Veräußerergruppe: Landeseigene Wohnungsunternehmen	19
Mieterkauf	99
Leerstand	81
Vermietet	38
Typische GFZ: 0,0	1
Typische GFZ: 0,1	11
Typische GFZ: 0,2	62
Typische GFZ: 0,8	34
Typische GFZ: 1,0	46
Typische GFZ: 1,5	8
Gute Wohnlage	194
Sehr gute Wohnlage	67
Typische Nutzungsart: sonstige	5

Es wiesen deutlich weniger als 5 % der Beobachtungen Auffälligkeiten auf, so dass die Beobachtungen im Datensatz verbleiben. Der von D'Agostino et al. (1990) vorgeschlagene Test auf Normalverteilung zeigt nicht normalverteilte Residuen an, so dass im folgenden eine robuste, stufenweise ML-Schätzung vorgenommen wird. Abbildung 8.4 auf Seite 82 zeigt den Residualplot.

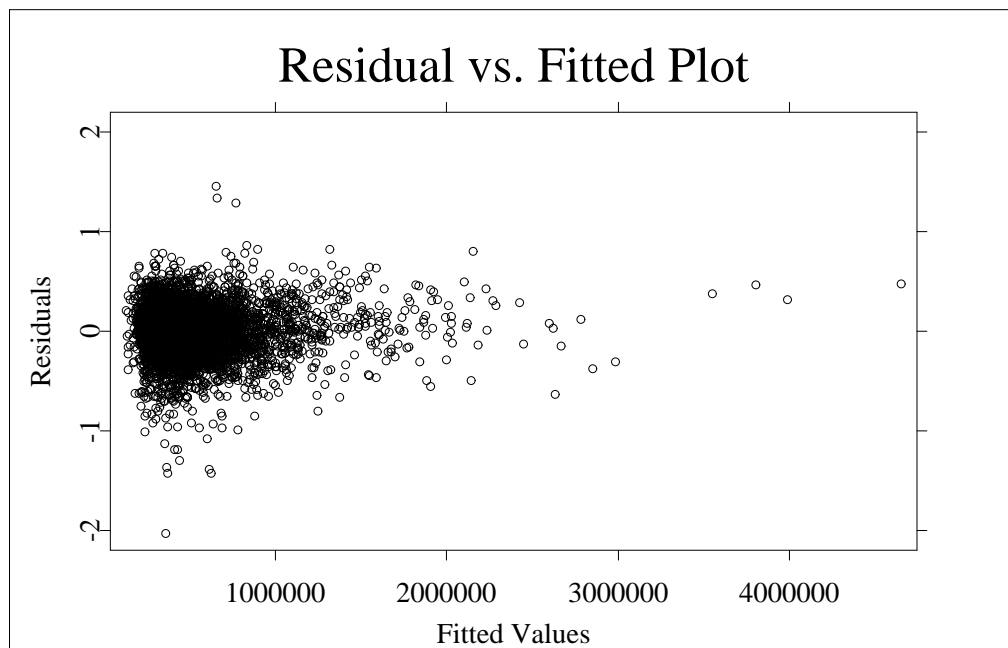


Abbildung 8.4: Residual vs. Fitted Plot für den Teilzeitraum 1996 - 2000

8.3.2 Regressionskennzahlen für die Teilregression

Aus Gründen der Vergleichbarkeit wurden auch in der Regression über den Teilzeitraum 1996 bis 2000 anstelle der Bezirks- und Lagevariablen zunächst Bezirksdummies gesetzt. Die Regressionskennzahlen finden sich in Tabelle 8.10.

Tabelle 8.10: *Regressionskennzahlen der Teilregression: Modell mit Bezirksdummies*

Number of Obs	6.235
F(48, 6.186)	291,91
Prob>F	0,0000
R^2	0,7369
\bar{R}^2	0,7354
Root MSE	0,2496

Ersetzt man die Bezirksdummies nur durch Bezirksvariablen und lässt die Lagevariablen erst einmal außen vor, ist das korrigierte Bestimmtheitsmaß ebenfalls geringer, als wenn man anstelle der Bezirksvariablen Bezirksdummies setzt, nämlich um 1,28 %. Damit werden wiederum die Bezirksdummies

weitestgehend vollständig durch die Bezirksvariablen erklärt, ohne dass sich jedoch das Modell an sich verbessert (s. Tabelle 8.11).

Tabelle 8.11: *Regressionskennzahlen der Teilregression: Modell mit Bezirksvariablen*

Number of Obs	6.235
F(41, 6.193)	318,40
Prob>F	0,0000
R^2	0,7275
\bar{R}^2	0,7261
Root MSE	0,2539

Zum Schluss wurde das volle Modell mit Lage- und Bezirksvariablen anstelle der Bezirksdummies geschätzt. Wie man Tabelle 8.12 entnehmen kann, steigt durch die Lagevariablen das korrigierte Bestimmtheitsmaß von 73,54 % auf 75,53 % oder um 2,71 %.

Tabelle 8.12: *Regressionskennzahlen der Teilregression: Modell mit Bezirks- und Lagevariablen*

Number of Obs	5.977
F(51, 5.925)	313,57
Prob>F	0,0000
R^2	0,7570
\bar{R}^2	0,7553
Root MSE	0,2420

Insgesamt wurden weniger Beobachtungen für die Regression benutzt als in den ersten beiden Fällen, da die Lagevariablen auch nach 1996 nicht vollständig erfasst wurden.

8.3.3 Koeffizienten der Teilregression: Hauseigenschaften

Die Koeffizienten der Hauseigenschaften können für die Teilregression Tabelle 8.13 auf Seite 84 entnommen werden. In diesem Teil wurden keine neuen Variablen in die Regression aufgenommen.

Tabelle 8.13: *Koeffizienten der Hauseigenschaften der Teilregression über den Zeitraum 1996 - 2000*

Variable	Coef.	t	P> t
Alter	-0,003	-16,08	0,000
ln(Grundstücksfläche)	0,341	23,34	0,000
ln(Geschossfläche)	0,407	25,47	0,000
Dachausbau	0,042	4,39	0,000
Geschäftliche Bindung	0,328	2,50	0,012
Ungewönl. Verhältnisse am Grundstück	-0,014	-1,70	0,089
Persönliche Verhältnisse des Geschäftsverkehrs	-0,214	-6,11	0,000
Ungew. Verhältn. am Grundstück tatsächl. Art	-0,082	-4,06	0,000
Wasserlage	0,363	8,06	0,000
Flachdach	-0,022	-2,01	0,045
Walmdach	0,065	6,56	0,000
Zeltdach	0,064	3,14	0,002
Erwerber: sonst. jur. Person	0,158	3,64	0,000
Zustand: gut	0,159	16,37	0,000
Zustand: schlecht	-0,189	-14,48	0,000
Doppelhaushälfte	-0,038	-3,79	0,000
Reihenhaus	0,032	1,96	0,050
Reihenendhaus	-0,048	-3,34	0,001
Zwei Vollgeschosse	0,057	5,46	0,000
Drei oder vier Vollgeschosse	0,091	2,45	0,014
Veräußerer: BRD	-0,150	-5,16	0,000
Veräußerer: Gemeinn. Wohnunt.	-0,133	-7,19	0,000
Veräußerer: sonst. jur. Person	0,044	2,92	0,003
Veräußerer: Landeseigene Wohnungsunt.	-0,194	-4,94	0,000
Mieterkauf	-0,124	-6,05	0,000
Leerstand	-0,117	-1,91	0,056
Nicht bezugsfertig	0,040	2,76	0,006
Vermietet	-0,157	-4,27	0,000

Anmerkung: Die abhängige Variable ist der Logarithmus des Preises. Die Schätzung erfolgte gem. Gleichung 8.1 stufenweise und robust nach der Maximum-Likelihood-Methode. Die Variablen sind im Anhang A.1 erklärt.

Vergleicht man die Koeffizienten in den Tabellen 8.13 und 8.5, fällt auf, dass sich zum einen die Höhe der Koeffizienten geändert hat, zum anderen aber auch einige Variablen insignifikant geworden sind, während vorher insignifikante Variablen jetzt teilweise signifikant sind.

Tests auf Gleichheit der Koeffizienten ergaben, dass die Koeffizienten für

$\ln(\text{Grundstücksfläche})$, $\ln(\text{Geschossfläche})$, Dachausbau, Zustand: gut, Doppelhaushälfte, vermietet und Veräußerer: Gemeinnützige Wohnungsunternehmen auf einem Signifikanzniveau von 5 % voneinander verschieden sind. Ein Anstieg der Grundstücksfläche um 1 % führt nun zu einem Preisanstieg von 0,341 % anstelle eines Preisanstiegs von 0,295 %, 1 % zusätzliche Geschossfläche zu einem Preisanstieg von 0,407 % statt 0,498 %. Der Koeffizient für das Alter ist dagegen bei beiden Regressionen gleich geblieben. Allerdings beträgt das durchschnittliche Alter jetzt 40,10 Jahre (s. Tabelle 8.14). Trotzdem ergibt sich weiterhin eine Preisänderung von 0,12 % nach unten, wenn das durchschnittliche Haus um 1 % altert.

Tabelle 8.14: *Deskriptive Statistiken für Alter, Grundfläche, Geschossfläche und Preis für den Teilzeitraum 1996 - 2000*

Variable	Obs	Mean	Median	Std. Dev.	Min	Max
Alter	6644	40,10	39,00	27,0208	0	200
Grundstücksfläche	6632	569,70	514,00	339,4162	100	7045
Geschossfläche	6644	147,30	136,00	65,7240	24	2885
Preis	6644	551.272	470.000	379.685	41.000	7.500.000

Die Variable **Ungewöhnliche Verhältnisse des Geschäftsverkehrs** ist jetzt insignifikant, dafür ist die Variable **Ungewöhnliche Verhältnisse am Grundstück** signifikant geworden. Liegen diese vor, nimmt der Kaufpreis um 1,4 % ab.

Die möglichen Blocklagen Frontgrundstück, Eckgrundstück, Mehrfrontengrundstück, Hammergrundstück und Hinterlandgrundstück sind jetzt alle insignifikant, während bei der Regression über den Gesamtzeitraum das Hammergrundstück zu einem Preisabschlag von 4,0 % führte. Anstelle des Pultdachs führt jetzt ein Flachdach zu einem Abschlag von 2,2 %.

Der Koeffizient für die Erwerbergruppe Gesellschaft bürgerlichen Rechts ist insignifikant geworden.

Zusätzlich zu den vorher signifikanten Gebäudestellungen Doppelhaushälfte und Reihenendhaus ist jetzt auch der Koeffizient für ein Reihenhaus signifikant. Handelt es sich um ein solches, so steigt der Preis um 3,2 % gegenüber dem eines Einzelhauses. Damit ist ein Reihenhaus beliebter als ein Einzelhaus, und dieses wiederum beliebter als ein Doppel- oder Reihenendhaus. Ein Grund für den Kauf eines Reihenhauses oder auch eines Doppel-

hauses gegenüber einem Einzelhaus sind die geringeren Kosten. Eigentlich wurde erwartet, dass ein Einzelhaus höher geschätzt würde, da nur dieses nach individuellen Vorstellungen errichtet werden kann und den Bewohnern den größten Freiraum bietet. Reihenhäuser sind dagegen "vorwiegend durch Bauträger errichtete, kostengünstige Familienheime, die mit relativ kleinem Grundstücksanteil auskommen und dennoch den vollen Komfort eines eigenen Heims bieten. Die Grundrisse sind weitgehend standardisiert" (Gerhards und Keller; 2002). Ein Reihenhaus könnte für den Käufer einen Kompromiss zwischen einer Wohnung und einem Einzelhaus darstellen. Während die enge Verbindung zur Nachbarsfamilie in einem Doppelhaus offensichtlich eher gescheut wird, scheint das Reihenhaus für größere Anonymität zu sorgen.

Bei der Verfügbarkeit zeigt sich, dass neben dem Mieterkauf und einem vermieteten Haus auch Leerstand zu einem Preisabschlag führt. Dieser beträgt 11,7 %. Ist das Haus dagegen nicht bezugsfertig, so beträgt der Preisaufschlag 4,0 %. Hierbei handelt es sich um Neubauten, die bereits vor der Fertigstellung gekauft werden.

8.3.4 Koeffizienten der Teilregression: Lageeigenschaften

Diese Lagevariablen wurden neu in die Regression aufgenommen und wurden bei der Regression über den gesamten Zeitraum auf Grund von missings nicht überprüft. Sie werden häuserindividuell erfasst und sind damit genauer als die Bezirksvariablen. Die Koeffizienten der Lagevariablen können Tabelle 8.15 entnommen werden.

Tabelle 8.15: *Koeffizienten der Lageeigenschaften der Teilregression über den Zeitraum 1996 - 2000*

Variable	Coef.	t	P> t
Typische Nutzungsart: sonstige	-0,242	-1,98	0,047
Einfache Wohnlage	-0,042	-4,97	0,000
Gute Wohnlage	0,111	11,96	0,000
Sehr gute Wohnlage	0,504	14,16	0,000
Typische GFZ: 0	0,381	2,21	0,027
Typische GFZ: 0,1	-0,346	-2,80	0,005
Typische GFZ: 0,2	-0,132	-5,14	0,000
Typische GFZ: 0,6	0,069	5,10	0,000
Typische GFZ: 0,8	-0,148	-3,49	0,000
Typische GFZ: 0,9	-0,234	-4,55	0,000

Fortsetzung auf der nächsten Seite

Tabelle 8.15: *Koeffizienten der Lageeigenschaften der Teilregression über den Zeitraum 1996 - 2000 - Fortsetzung*

Variable	Coef.	t	P> t
Typische GFZ: 1	0,102	2,94	0,003
Typische GFZ: 1,5	0,049	6,57	0,000
Typische GFZ: 4	0,191	4,04	0,000
<i>Anmerkung:</i> Die abhängige Variable ist der Logarithmus des Preises. Die Schätzung erfolgte gem. Gleichung 8.1 stufenweise und robust nach der Maximum-Likelihood-Methode. Die Variablen sind im Anhang A.1 erklärt.			

Liegt das Objekt in einer Gegend, die nicht nur Wohnbebauung, sondern auch eine andere Nutzung vorsieht, nämlich Dienstleistung, produzierendes Gewerbe oder Mischcharakter, so beträgt der Preisabschlag 24,2 %. Dies könnte beispielsweise an Lärm- oder Geruchsbelästigungen liegen, die von diesen Betrieben ausgehen können.

Während eine einfache Wohnlage nur zu einem Preisabschlag von 4,2 % gegenüber der mittleren Wohnlage führt, wird für ein bebautes Grundstück in guter Wohnlage 11,1 % mehr gezahlt, für ein Haus in sehr guter Lage sogar 50,4 %.

Die Referenzausprägung für die **Typische Geschossflächenzahl** ist eine GFZ von 0,4. Eine GFZ von Null würde bedeuten, dass hier überhaupt keine Bebauung vorzufinden ist. Eine typische GFZ von weniger als 0,9 führt, mit Ausnahme einer typischen GFZ von 0,6, zu Preisabschlägen. Ab einer typischen GFZ von 1,0 steigt der Preis. Hierbei handelt es sich um Gegenden mit einer entsprechend dichten Bebauung.

8.3.5 Koeffizienten der Teilregression: Bezirkseigenschaften

Tabelle 8.16 enthält die Koeffizienten der Bezirkseigenschaften.

Tabelle 8.16: *Koeffizienten der Bezirkseigenschaften der Teilregression über den Zeitraum 1996 - 2000*

Variable	Coef.	t	P> t
ln(Baum)	-0,053	-2,58	0,010
Anzahl U-Bahnhöfe pro 10.000 Einwohner	0,224	7,70	0,000

Fortsetzung auf der nächsten Seite

Tabelle 8.16: *Koeffizienten der Bezirkseigenschaften der Teilregression über den Zeitraum 1996 - 2000 - Fortsetzung*

Variable	Coef.	t	P> t
ln(Bevölkerung unter 6/Grundschulklasse)	0,144	6,44	0,000
ln(Spielfläche/Kind unter 15)	-0,057	-3,97	0,000
ln(Anteil Kinder)	-0,337	-6,04	0,000
ln(Arbeitslose/Erwerbsperson)	-0,287	-7,61	0,000
ln(Haushaltsnettoeinkommen)	0,639	8,33	0,000
ln(Ausländer/Einwohner)	0,177	12,43	0,000

Anmerkung: Die abhängige Variable ist der Logarithmus des Preises. Die Schätzung erfolgte gem. Gleichung 8.1 stufenweise und robust nach der Maximum-Likelihood-Methode. Die Variablen sind im Anhang A.1 erklärt.

Die Koeffizienten der Bezirkseigenschaften müssen zusammen mit den Lagevariablen betrachtet werden.

Vergleicht man diese Variablen mit denen der Tabelle 8.16, so fällt auf, dass die Variablen **Anteil Waldfläche**, **Anzahl S-Bahnhöfe pro 100 Hektar** und **Nichterwerbspersonen/Erwerbsperson** nicht mehr signifikant sind. Insbesondere der **Anteil Waldfläche** und die **Anzahl S-Bahnhöfe pro 100 Hektar** scheinen durch die Variable **Wohnlage** erfasst zu werden. Gute und sehr gute Wohnlagen sind durch "Baumbepflanzung an der Straße oder Gärten" (Gerhards und Keller; 2002) und "günstige Verkehrsanbindung auch mit öffentlichen Verkehrsmitteln" (Gerhards und Keller; 2002) gekennzeichnet. Trotzdem bleibt die Variable **Anzahl U-Bahnhöfe pro 10.000 Einwohner** weiter signifikant. Das könnte daran liegen, dass die S-Bahn dasjenige Verkehrsmittel ist, das die sehr guten Wohnlagen Zehlendorf und Wilmersdorf sowie die guten Wohnlagen Steglitz und Köpenick mit dem Zentrum verbindet, während die U-Bahn ein Verkehrsmittel ist, das vorwiegend im Zentrum verkehrt. Weiterhin fällt auf, dass der Koeffizient für **ln(Baum)** negativ geworden ist. Nimmt die Anzahl der Bäume pro 1 km Straßenlänge um 1 % zu, fällt der Preis eines bebauten Grundstücks um 0,053 %. Dies widerspricht der Charakterisierung der guten und sehr guten Wohnlagen.

8.3.6 Koeffizienten der Teilregression: Jahresdummies

Die Koeffizienten der Jahresdummies können Tabelle 8.17 auf Seite 89 entnommen werden.

Tabelle 8.17: *Koeffizienten der Jahresdummies der Teilregression über den Zeitraum 1996 - 2000*

Variable	Coef.	t	P> t
1999	-0,042	-4,92	0,000
2000	-0,091	-8,70	0,000
Konstante	2,814	4,78	0,000

Anmerkung: Die abhängige Variable ist der Logarithmus des Preises. Die Schätzung erfolgte gem. Gleichung 8.1 stufenweise und robust nach der Maximum-Likelihood-Methode. Die Variablen sind im Anhang A.1 erklärt.

Das Jahr 1996 ist hier der Referenzzustand. Die Koeffizienten für die Jahre 1997 und 1998 sind insignifikant. Die Koeffizienten für 1999 und 2000 deuten auf fallende Preise gegenüber 1996 hin, was sich mit den Ergebnissen der Regression über den Gesamtzeitraum deckt.

8.3.7 Vergleich der Regressionen

Um die Aussagekraft der verschiedenen Regressionen zu vergleichen, werden in Tabelle 8.18 noch einmal die korrigierten Bestimmtheitsmaße der Regressionen zusammengefasst.

Tabelle 8.18: *Korrigierte Bestimmtheitsmaße für die verschiedenen Regressionen*

Gesamtzeitraum und Bezirksdummies	74,25 %
Gesamtzeitraum und Bezirksvariablen	73,69 %
Teilzeitraum und Bezirksdummies	73,54 %
Teilzeitraum und Bezirksvariablen	72,61 %
Teilzeitraum und Bezirks- sowie Lagevariablen	75,53 %

Die Bezirksvariablen ersetzen weitgehend die Bezirksdummies und schaffen es so, die Preisunterschiede in den einzelnen Bezirken in ihre Bestandteile zu zerlegen und zu erklären. Dies geht geringfügig auf Kosten der Erklärungskraft des Modells.

Im Gegensatz zu den Bezirksvariablen werden die Lagevariablen für jedes Haus individuell erfasst. Sie beschreiben also die unmittelbare Umgebung des betrachteten Objekts und sind auf diese Weise präziser als die Bezirksvariablen. Es zeigt sich, dass einige Bezirksvariablen dasselbe messen wie die Lagevariablen, nur auf einer gröberen Ebene. Die häuserindividuelle Schätzung führt dabei zu einem besseren Ergebnis.

Kapitel 9

Zusammenfassung und Ausblick

Das Ziel dieser Arbeit, eine hedonische Preisfunktion zu ermitteln und Bezirkseigenschaften zu identifizieren, die die Attraktivität der Bezirke bestimmen, wurde erreicht. Es konnten Variablen gefunden werden, die die Preisunterschiede zwischen den einzelnen Bezirken anhand des Anteils an Grünflächen, der Verkehrs- und sozialen Infrastruktur, der sozio-ökonomischen Lage, der Altersstruktur und des Ausländeranteils messen. Dabei hat sich gezeigt, dass die Bezirksvariablen zwar die Preisunterschiede zwischen den einzelnen Bezirken in ihre Bestandteile zerlegen können, aber zu unpräzise sind, um die Erklärungskraft des Modells zu erhöhen. Das wird erst mit den Lagevariablen, die für jedes Haus individuell erfasst werden, erreicht.

In Zukunft sollte also versucht werden, schon bei der Erfassung des Datensatzes individuelle Lageparameter vollständig zu erheben, auch bisher nicht betrachtete wie z. B. Lärm- oder Geruchsbelästigung. Es könnte versucht werden, die Bezirksvariablen zu individualisieren, indem beispielsweise die Entfernung zur nächsten U- oder S-Bahnstation oder zur nächsten Schule in Metern angegeben wird. Da Einfamilienhausgebiete häufig nicht über den gesamten Bezirk verteilt sind, sondern sich auf einzelne Ortsteile beschränken, wäre es auch nützlich, die bisher betrachteten Bezirksvariablen zumindest auf Ortsteilebene zu erheben und zu analysieren.

Verschiedene Variablen konnten aus Mangel an Daten nicht betrachtet werden. Dazu gehörten z. B. die Kriminalitätsrate und die Ladenflächen im Bezirk, wobei sich für letztere auch eine individualisierte Angabe in Metern zum nächsten Laden anbietet. Es ist zu befürchten, dass die Bezirksfusion von 23 auf zwölf Bezirke dazu führt, dass in Zukunft aus Kostengründen die Erhebung von Daten auf einer noch gröberen Ebene erfolgt. Daher besteht die Notwendigkeit, der Politik die Wichtigkeit umfassender statistischer

Datenerhebungen zu verdeutlichen, da statistische Analysen dazu beitragen können, wirtschaftliche Vorgänge besser zu verstehen und zu erklären.

Da nicht in jedem Bezirk Einfamilienhäuser verkauft wurden, führte dies eventuell bei der Variable "Ausländeranteil an der Bevölkerung" zu Verzerrungen. Um den Einfluss von Bezirksvariablen auf das gesamte Berliner Gebiet untersuchen zu können, könnten beispielsweise die Märkte für Eigentums- oder Mietwohnungen untersucht werden, da sich diese wahrscheinlich weniger auf einzelne Ortsteile beschränken. Auch könnte ein Vergleich der Einflussfaktoren zwischen verschiedenen deutschen Städten, zwischen Groß- und Kleinstädten oder auch international erfolgen.

Anhang A

Liste der Variablen

Dieser Anhang ist dreiteilig aufgebaut: in Tabelle A.1 finden sich die wichtigsten Hausvariablen und deren Beschreibung, in Tabelle A.2 die Lagevariablen, die im Gegensatz zu den Bezirksvariablen häuserindividuell erfasst werden, und in Tabelle A.3 schließlich die Bezirksvariablen und deren Erklärung.

A.1 Die Hausvariablen und deren Erklärung

Tabelle A.1: *Liste der Hausvariablen*

Variable	Erklärung
Metrische Variablen	
Preis	Gibt den Kaufpreis in DM an.
Alter	Aus dem Baujahr des Hauses und dem Vertragsdatum generierte Variable. Sie gibt das Alter des Hauses in Jahren an.
Grundstücksfläche	Gibt die Grundstücksfläche in Quadratmetern an.

Fortsetzung auf der nächsten Seite

Tabelle A.1: *Liste der Hausvariablen - Fortsetzung*

Variable			Erklärung
Anzahl	der	Vollge-	Nach §20 Abs. 1 BauNVO gelten als "Vollgeschosse [...] Geschosse, die nach landesrechtlichen Vorschriften Vollgeschosse sind oder auf ihre Zahl angerechnet werden." Die dazugehörige landesrechtliche Vorschrift für Berlin findet sich in §2 Abs. 4 BauO Bln. Diese regelt folgendes: "Vollgeschosse sind Geschosse, deren Deckenoberkante im Mittel mehr als 1,40 m über die festgelegte Geländeoberfläche hinausragt und die über mindestens zwei Drittel ihrer Grundfläche eine lichte Höhe von mindestens 2,30 m haben. Ein gegenüber den Außenwänden des Gebäudes zurückgesetztes oberstes Geschoß (Staffelgeschoß) und Geschosse im Dachraum sind nur dann Vollgeschosse, wenn sie die lichte Höhe gemäß Satz 1 über mindestens zwei Drittel der Grundfläche des darunterliegenden Geschosses haben."
schosse			
		Geschossfläche	Nach §20 Abs. 3 BauNVO ist die "Geschoßfläche nach den Außenmaßen der Gebäude in allen Vollgeschossen zu ermitteln. Im Bebauungsplan kann festgesetzt werden, daß die Flächen von Aufenthaltsräumen in anderen Geschossen einschließlich der zu ihnen gehörenden Treppenträume und einschließlich ihrer Umfassungswände ganz oder teilweise mitzurechnen oder ausnahmsweise nicht mitzurechnen sind." §20 Abs. 4 BauNVO regelt, dass bei "der Ermittlung der Geschoßfläche [...] Nebenanlagen [...], Balkone, Loggien, Terrassen sowie bauliche Anlagen, soweit sie nach Landesrecht in den Abstandsflächen (seitlicher Grenzabstand und sonstige Abstandsflächen) zulässig sind oder zugelassen werden können, unberücksichtigt" bleiben.

Fortsetzung auf der nächsten Seite

Tabelle A.1: *Liste der Hausvariablen - Fortsetzung*

Variable	Erklärung
Ungewöhnliche Verhältnisse	
Persönliche Verhältnisse des Geschäftsverkehrs	Persönliche Verhältnisse des Geschäftsverkehrs liegen dann vor, wenn die Veräußerung zum Zwecke der Erbaueinandersetzung erfolgte oder besonderen Bindungen verwandtschaftlicher, wirtschaftlicher oder sonstiger Art zwischen den Vertragsparteien unterlag. Dieser Eintrag findet sich, wenn der Erwerber der Nachbar ist, bei Arrondierung, Erbteilung, Verkauf unter Verwandten oder Scheidungsaueinandersetzungen (vgl. Sofyan et al. (2002)).
Ungewöhnliche Verhältnisse am Grundstück	Diese liegen dann vor, wenn in Abteilung II des Grundbuchs Lasten oder Beschränkungen eingetragen sind (vgl. Sofyan et al. (2002)). Dabei handelt es sich um Rechte Dritter bzw. anderer Grundstücke an dem belasteten Grundstück. Grunddienstbarkeiten stehen zwingend, Reallasten und Vorkaufsrechte fakultativ dem jeweiligen Eigentümer eines anderen Grundstücks zu (vgl. Lippe et al. (1994)).
<ul style="list-style-type: none"> • Grunddienstbarkeit (§§1018 ff. BGB): Eintragung eines der folgenden Rechte: <ul style="list-style-type: none"> – "bestimmte Benutzung (Wegerecht, Brunnenrecht) – Verbot bestimmter Handlungen (Wohnverbot, Verbot bestimmter Gewerbebetriebe, z. B. wegen Lärmbelästigung) – Unterlassen bestimmter Einwirkungen auf das andere Grundstück (Schatten durch Bäume, Rauchbelästigung) – Duldung bestimmter Einwirkungen des anderen Grundstücks" (Lippe et al.; 1994). 	

Fortsetzung auf der nächsten Seite

Tabelle A.1: *Liste der Hausvariablen - Fortsetzung*

Variable	Erklärung
<i>- Fortsetzung -</i>	
	<ul style="list-style-type: none"> • Erbbaurecht: "das vererbbare und veräußerbare Recht, auf dem Grundstück ein Gebäude zu errichten und zu unterhalten" (Lippe et al.; 1994). • Nießbrauchrecht (§§1030 ff. BGB): "das (unübertragbare) Recht, die regelmäßigen Nutzungen (Erträge) eines Grundstücks zu ziehen [...], z. B. Obsternte, Mieten" (Lippe et al.; 1994). • Vorkaufsrecht (§§504 ff. BGB, §§1094 ff. BGB): "die Berechtigung des Begünstigten [...], in einen vom Eigentümer mit einem Dritten geschlossenen Kaufvertrag über das Grundstück zu denselben Bedingungen einzutreten" (Lippe et al.; 1994). • Reallast (§§1105 BGB): "wiederkehrende Leistungen aus dem Grundstück an den Begünstigten (Geld, Naturalien; Erntehilfe)" (Lippe et al.; 1994).
Ungewöhnliche Verhältnisse des Geschäftsverkehrs	Ver- Ge- Hierunter fällt z. B. die Begleichung des Kaufpreises durch Raten- oder Rentenzahlungen bzw. Einräumung eines Rechtes, Leibrenten oder die Übernahme von Hypotheken (vgl. Sofyan et al. (2002)).
Ungewöhnliche Verhältnisse tatsächlicher Art	Ver- Hierunter fallen ungewöhnliche Einzelfälle.

Fortsetzung auf der nächsten Seite

Tabelle A.1: *Liste der Hausvariablen - Fortsetzung*

Variable	Erklärung
Kategorielle Variablen	
Blocklage	Es kommen folgende Blocklagen vor: Frontgrundstück, Eckgrundstück, Mehrfrontengrundstück, Hammergrundstück und Hinterlandgrundstück.
Dachform	<p>Es traten folgende Dachformen auf: Flachdach, Pultdach, Satteldach, Berliner Dach, Walmdach, Mansardendach, Zeltdach und Sheddach. Im folgenden sind einige Dachformen beschrieben.</p> <ul style="list-style-type: none"> • Flachdach: Dachform mit nur sehr geringer Dachneigung, die nur dazu dient, dass Regenwasser abfließen kann. • Pultdach: Dachform mit nur einer schrägen Dachfläche. • Satteldach: "häufig anzutreffende Dachform mit zwei gegeneinander geneigten Dachflächen, die von den Traufseiten ansteigend in einer bestimmten Dachneigung im First zusammentreffen. Die Dachneigung entscheidet meist darüber, ob das Dachgeschoss ausbaufähig ist" (Gerhards und Keller; 2002). • Walmdach: "Dachform, bei der die Giebelseiten als Dachfläche ausgebildet sind und die Seitenteile ähnlich dem Satteldach ansteigen" (Gerhards und Keller; 2002). • Mansardendach: Dachform, bei der der untere Teil steiler aufsteigt als der obere. Dadurch ist eine bessere Ausnutzung beim Dachgeschossausbau möglich.

Fortsetzung auf der nächsten Seite

Tabelle A.1: *Liste der Hausvariablen - Fortsetzung*

Variable	Erklärung
	- <i>Fortsetzung</i> -
	<ul style="list-style-type: none"> • Zeltdach: "Dachkonstruktion, die einem Zelt nachempfunden ist. Dabei sind alle vier Dachseitenteile gleich groß. Das Dach läuft spitz zusammen" (Gerhards und Keller; 2002). • Sheddach: Kombination mehrerer Pultdächer, besonders im Industriebau anzutreffen.
Erwerbergruppe	Es traten folgende Erwerbergruppen auf: Privat, Gesellschaft bürgerlichen Rechts, sonstige juristische Person und sonstige Erwerbergruppe. Hierunter wurden die Erwerbergruppen Bundesrepublik Deutschland, Land Berlin, sonstige öffentliche Hand, (gemeinnützige) Wohnungsunternehmen, Verfahrensträger, Versicherungsunternehmen/sonstige Fondsgesellschaft, Kirche/Religionsgemeinschaft und Diplomatische Vertretung zusammengefasst.
Gebäudestellung	Hier wurde unterschieden in Freistehend/Einzelhaus, Doppelhaushälfte, Reihenhaushaus und Reihenendhaus.
Veräußerergruppe	Folgende Veräußerergruppen gab es: Privat, Gesellschaft bürgerlichen Rechts, Bundesrepublik Deutschland, Land Berlin, sonstige öffentliche Hand, (gemeinnützige) Wohnungsunternehmen, sonstige juristische Person, Kirche/Religionsgemeinschaft und Landeseigene Wohnungsunternehmen. Unter sonstige Veräußerergruppe sind Immobilienfonds, Versicherungsunternehmen/sonstige Fondsgesellschaft, Diplomatische Vertretung und sonstiges Bundesland zusammengefasst.
Verfügbarkeit	Es wurde unterschieden in: verkäuferbewohnt, vermietet, Mieterkauf/Pächterkauf, Nutzungsrecht, Leerstand und nicht bezugsfertig.

Fortsetzung auf der nächsten Seite

Tabelle A.1: *Liste der Hausvariablen - Fortsetzung*

Variable	Erklärung
Zustandsnote	Dafür gab es drei mögliche Ausprägungen: normal, gut, schlecht.

A.2 Die Lagevariablen und deren Erklärung

Tabelle A.2: *Liste der Bezirksvariablen*

Variable	Erklärung
Wohnlage	<p>"Einfache Wohnlagen. Eine einfache Wohnlage ist gegeben, wenn das Wohnen z. B. durch Geräuschbelästigung oder aufgrund anderer Kriterien kontinuierlich erheblich beeinträchtigt wird und dadurch der Wohnwert gemindert wird. Mittlere Wohnlagen. a) Standard. Bei mittleren Wohnlagen handelt es sich um normale Wohnlagen ohne besondere Vor- und Nachteile. Die überwiegende Zahl der Wohnungen innerhalb des Stadtgebiets liegt in diesen Wohngebieten. Solche Wohngebiete sind zumeist dicht bebaut und weisen keine kontinuierlich beeinträchtigenden Belastungen durch Geräusch oder Geruch auf. b) Gut. Die guten Wohnlagen sind gekennzeichnet durch lockere Bebauung, Baumbepflanzung an der Straße oder Gärten, fehlenden Durchgangsverkehr, gute Einkaufsmöglichkeiten, nicht beeinträchtigende Einrichtungen und günstige Verkehrsanbindung auch mit öffentlichen Verkehrsmitteln zum Zentrum. [...] Sehr gute Wohnlagen. Die sehr guten Wohnlagen sind durch aufgelockerte, i. d. R. zweigeschossige, in Villenlagen bis zu viergeschossige Bebauung in ruhiger und verkehrsgünstiger Grünlage gekennzeichnet" (Gerhards und Keller; 2002).</p>

Fortsetzung auf der nächsten Seite

Tabelle A.2: *Liste der Lagevariablen - Fortsetzung*

Variable	Erklärung
Typische Geschossflächenzahl	Die Geschossflächenzahl wird in §20 BauNVO definiert und berechnet sich aus

$$\frac{\text{Geschossfläche}}{\text{Grundstücksfläche}}$$

Sie ist ein Maß baulicher Nutzung. Ihre zulässige Höhe wird im Bebauungsplan festgelegt. §17 BauNVO legt die Obergrenzen fest (s. Tabelle A.2).

Tabelle A.2: *Obergrenzen für die Bestimmung des Maßes baulicher Nutzung nach §17 BauNVO*

Baugebiet	GFZ
Kleinsiedlungsgebiete	0,4
reine Wohngebiete	1,2
allgemeine Wohngebiete	1,2
Ferienhausgebiete	1,2
besondere Wohngebiete	1,6
Dorfgebiete	1,2
Mischgebiete	3,0
Kerngebiete	3,0
Gewerbegebiete	2,4
Industriegebiete	2,4
sonstige Sondergebiete	2,4
Wochenendhausgebiete	0,2

Fortsetzung auf der nächsten Seite

Tabelle A.2: *Liste der Lagevariablen - Fortsetzung*

Variable	Erklärung
Typische Nutzungsart	<p>§1 Abs. 2 BauNVO unterscheidet u. a. folgende Baugebiete:</p> <ul style="list-style-type: none"> • Reine Wohngebiete dienen nach §3 BauNVO dem Wohnen. "Zulässig sind Wohngebäude. Ausnahmsweise können zugelassen werden <ol style="list-style-type: none"> 1. Läden und nicht störende Handwerksbetriebe, die zur Deckung des täglichen Bedarfs für die Bewohner des Gebiets dienen sowie kleine Betriebe des Beherbergungsgewerbes, 2. Anlagen für soziale Zwecke sowie den Bedürfnissen der Bewohner des Gebiets dienende Anlagen für kirchliche, kulturelle, gesundheitliche und sportliche Zwecke." <p>(§3 Abs. 2 und 3 BauNVO)</p> <ul style="list-style-type: none"> • §4 Abs. 1 - 3 BauNVO legt folgendes fest: "Allgemeine Wohngebiete dienen vorwiegend dem Wohnen. Zulässig sind <ol style="list-style-type: none"> 1. Wohngebäude, 2. die der Versorgung des Gebiets dienenden Läden, Schank- und Speisewirtschaften sowie nicht störende Handwerksbetriebe, 3. Anlagen für kirchliche, kulturelle, soziale, gesundheitliche und sportliche Zwecke. <p>Ausnahmsweise können zugelassen werden</p> <ol style="list-style-type: none"> 1. Betriebe des Beherbergungsgewerbes, 2. sonstige nicht störende Gewerbebetriebe, 3. Anlagen für Verwaltungen, 4. Gartenbaubetriebe, 5. Tankstellen."

Tabelle A.2: *Liste der Lagevariablen - Fortsetzung*

Variable	Erklärung
	- <i>Fortsetzung</i> -
	<ul style="list-style-type: none"> • "Mischgebiete dienen dem Wohnen und der Unterbringung von Gewerbebetrieben, die das Wohnen nicht wesentlich stören. Zulässig sind <ol style="list-style-type: none"> 1. Wohngebäude, 2. Geschäfts- und Bürogebäude, 3. Einzelhandelsbetriebe, Schank- und Speisewirtschaften sowie Betriebe des Beherbergungsgewerbes, 4. sonstige Gewerbebetriebe, 5. Anlagen für Verwaltungen sowie für kirchliche, kulturelle, soziale, gesundheitliche und sportliche Zwecke, 6. Gartenbaubetriebe, 7. Tankstellen, 8. Vergnügungsstätten [...]" <p>(§6 Abs. 1 und 2 BauNVO).</p> <ul style="list-style-type: none"> • "Gewerbegebiete dienen vorwiegend der Unterbringung von nicht erheblich belästigenden Gewerbebetrieben. Zulässig sind <ol style="list-style-type: none"> 1. Gewerbebetriebe aller Art, Lagerhäuser, Lagerplätze und öffentliche Betriebe, 2. Geschäfts-, Büro- und Verwaltungsgebäude, 3. Tankstellen, 4. Anlagen für sportliche Zwecke.

Fortsetzung auf der nächsten Seite

Tabelle A.2: *Liste der Lagevariablen - Fortsetzung*

Variable	Erklärung
	- <i>Fortsetzung</i> -
	Ausnahmsweise können zugelassen werden
	1. Wohnungen für Aufsichts- und Bereitschaftspersonen sowie für Betriebsinhaber und Betriebsleiter, die dem Gewerbebetrieb zugeordnet und ihm gegenüber in Grundfläche und Baumasse untergeordnet sind,
	2. Anlagen für kirchliche, kulturelle, soziale und gesundheitliche Zwecke,
	3. Vergnügungsstätten.”
	(§8 Abs. 1 - 3 BauNVO)

A.3 Die Bezirksvariablen und deren Erklärung

Tabelle A.3: *Liste der Bezirksvariablen*

Variable	Erklärung
Anteil Waldfläche an der Bezirksfläche	Aus den Variablen Waldfläche und Bezirksfläche generierte Variable, die angibt, welcher Anteil der Bezirksfläche mit Sträuchern oder Bäumen bewachsen ist.

Fortsetzung auf der nächsten Seite

Tabelle A.3: *Liste der Bezirksvariablen - Fortsetzung*

Variable	Erklärung
Anzahl der Nichterwerbspersonen auf eine Erwerbsperson	<p>”Erwerbspersonen sind Personen, die mindestens im Alter von 15 Jahren sind und die eine auf den Erwerb ausgerichtete Tätigkeit ausüben oder suchen - also Erwerbstätige und Erwerbslose. Nach dem Inländerkonzept müssen die Personen des weiteren ihren Wohnsitz im betrachteten Bundesland haben. Der Ertrag der Tätigkeit und sein Beitrag zum Lebensunterhalt sowie die vertragsmäßige oder tatsächliche Arbeitszeit sind für die Zuordnung zu dem Kreis der Erwerbspersonen nicht von Bedeutung” (Statistisches Landesamt Berlin; 2000). Da die Zahl der Erwerbspersonen nicht auf Bezirksebene erfasst wird, wurde diese Variable aus der fortgeschriebenen Bevölkerung nach Bezirken und Altersgruppen generiert:</p> $\frac{\text{Bevölkerung unter 15} + \text{Bevölkerung älter als 65}}{\text{Bevölkerung im Alter von 15 bis unter 65 Jahren}}$ <p>Dieser Quotient gibt an, wie viele Nichterwerbspersonen durch eine Erwerbsperson unterhalten werden.</p>

Fortsetzung auf der nächsten Seite

Tabelle A.3: *Liste der Bezirksvariablen - Fortsetzung*

Variable	Erklärung
Anteil der Arbeitslosen an der Erwerbsbevölkerung	<p>Diese Variable wurde aus der Anzahl der Arbeitslosen eines Bezirks und der fortgeschriebenen Bevölkerung nach Bezirken und Altersgruppen generiert:</p> $\frac{\text{Arbeitslose}}{\text{Bevölkerung im Alter von 15 bis unter 65 Jahren}}$ <p>”Als arbeitslos in einer Region gelten Personen, die in dieser Region wohnen, vorübergehend in keinem Beschäftigungsverhältnis stehen und sich beim Arbeitsamt gemeldet haben, um in ein Arbeitsverhältnis als Arbeitnehmer/-innen oder Heimarbeiter/-innen vermittelt zu werden” (Statistisches Landesamt Berlin; 2000). Diese Zahl ist nicht zu verwechseln mit der offiziellen Arbeitslosenquote, die aus</p> $\frac{\text{Arbeitslose}}{\text{Arbeitskräftepotential}}$ <p>gebildet wird.</p>
Anteil der Ausländer an der Bevölkerung	Ausländer sind ”Personen mit ausschließlich ausländischer oder ungeklärter Staatsangehörigkeit sowie Staatenlose” (Statistisches Landesamt Berlin; 2000). Die Zahl der Ausländer wurde in Bezug gesetzt zur fortgeschriebenen Bevölkerung nach Bezirken.
Anteil der Kinder an der Bevölkerung	<p>Aus der Bevölkerung nach Altersgruppen generierte Variable:</p> $\frac{\text{Bevölkerung unter 6}}{\text{Gesamtbevölkerung}} + \frac{\text{Bevölkerung von 6 bis unter 15}}{\text{Gesamtbevölkerung}}$
Baum	Anzahl der Bäume auf 1 km Straßenlänge.

Fortsetzung auf der nächsten Seite

Tabelle A.3: *Liste der Bezirksvariablen - Fortsetzung*

Variable	Erklärung
Bevölkerung	"Zur Wohnbevölkerung eines Ortes zählten alle Einwohner, die dort ihren Lebensmittelpunkt hatten. Als Lebensmittelpunkt wurde der Ort gewertet, von dem aus der Einwohner zur Arbeit oder Ausbildung ging bzw. an dem er sich überwiegend aufhielt, unabhängig davon, ob er - im Falle mehrerer Wohnungen - dort seine Haupt- oder Nebenwohnung hatte" (Statistisches Landesamt Berlin; 2000). Die Bevölkerung wird nach folgenden Altersgruppen ausgewiesen: Bevölkerung unter 6 Jahren, Bevölkerung im Alter von 6 bis unter 15 Jahren, Bevölkerung im Alter von 15 bis unter 25 Jahren, Bevölkerung im Alter von 25 bis unter 45 Jahren, Bevölkerung im Alter von 45 bis unter 65 Jahren, Bevölkerung im Alter von 65 und älter.
Nettoeinkommen	"Das monatliche Nettoeinkommen ergibt sich für die abhängigen Erwerbstätigen aus den Bruttoeinkommen im Monat abzüglich Steuern und Sozialversicherung" (Statistisches Landesamt Berlin; 2000). "Einen Privathaushalt bilden alle Personen, die gemeinsam wohnen und wirtschaften, d. h. insbesondere ihren Lebensunterhalt gemeinsam finanzieren. [...] Wer allein wirtschaftet, bildet einen eigenen Haushalt, und zwar auch dann, wenn er mit anderen Personen eine gemeinsame Wohnung hat (z. B. Mitglieder von Wohngemeinschaften, Untermieter)" (Statistisches Landesamt Berlin; 2000).
Kinder unter 6 Jahren pro Grundschulklasse	Dieser Quotient gibt die Anzahl der Kinder unter 6 Jahren pro Grundschulklasse an. Er wurde gebildet aus <div style="text-align: center;"> $\frac{\text{Bevölkerung unter 6 Jahren}}{\text{Anzahl Grundschulklassen}}$ </div>

Fortsetzung auf der nächsten Seite

Tabelle A.3: *Liste der Bezirksvariablen - Fortsetzung*

Variable	Erklärung
Spielfläche pro Bevölkerung unter 15	<p>Öffentliche Kinderspielflächen werden entsprechend den Spielbedürfnissen der verschiedenen Altersgruppen in unterschiedlichen Nutzungsarten angelegt. Die Pflege und Unterhaltung dieser Plätze erfolgt überwiegend durch die Gartenämter der Bezirke.</p> <ul style="list-style-type: none"> • Kleinkinderspielflächen: geeignet für Kinder unter sechs Jahren. Das Spielplatzangebot besteht bei einer nutzbaren Spielfläche von mindestens 150 m^2 vorwiegend aus einem Sandkasten, einer Rutsche und einer Schaukel. • Allgemeine Spielplätze: für Kinder und Jugendliche mit einer nutzbaren Spielfläche von mindestens 2.000 m^2; hier sind auch Spielangebote für Kleinkinder integriert. Für die über Sechsjährigen ist das Spielangebot durch unterschiedlichste Spielgeräte (z. B. Kletterkombinationen, Drehscheiben, Seilbahn) erweitert. Ballspielmöglichkeiten, Abenteuerspielecken und Wasserspielbereiche können eingeordnet sein. • Pädagogisch betreute Spielplätze für Kinder ab sechs Jahren mit einer nutzbaren Spielfläche von möglichst 4.000 m^2. Zu dieser Nutzungsart zählen Abenteuerspielplätze, Bauspielplätze und Kinderbauernhöfe. Das Angebot der pädagogisch betreuten Spielplätze steht nur an Wochentagen und während der Arbeitszeit der Betreuer zur Verfügung. Pflege, Wartung und Betreuung dieser Plätze erfolgt durch die Jugendeinrichtungen der Bezirke.

Fortsetzung auf der nächsten Seite

Tabelle A.3: *Liste der Bezirksvariablen - Fortsetzung*

Variable	Erklärung
<i>- Fortsetzung -</i>	
	<ul style="list-style-type: none"> • Spielplätze mit zeitlich begrenzter Nutzungsmöglichkeit für Kinder über sechs Jahre. Hierbei handelt es sich um die zusätzliche Nutzung von Spielflächen auf Schulhöfen außerhalb der Schulzeit. • Waldspielplätze: in der Regel für Kinder ab sechs Jahre. Die naturbelassenen Spielelemente wurden vorwiegend aus Holz der Berliner Wälder gefertigt. Die Pflege und Wartung der Waldspielfläche erfolgt durch die Berliner Forsten."
	<p>(http://www.stadtentwicklung.berlin.de). Nach §4 Abs. 1 Kinderspielplatzgesetz gilt für die Bemessung des Bedarfs an öffentlicher Spielplatzfläche ein Richtwert von 1,0 m^2 nutzbarer Fläche je Einwohner. Da Spielplätze normalerweise für Kinder bis 14 Jahre vorgesehen sind, wurde die Spielfläche näherungsweise zur Bevölkerung im Alter bis unter 15 Jahre in Beziehung gesetzt:</p> $\frac{\text{Spielfläche in } m^2}{\text{Bevölkerung unter 15 Jahren}}.$

Anhang B

Beobachtungen nach Bezirk und Jahr

Tabelle B.1: *Relative und absolute Häufigkeiten der Beobachtungen nach Bezirk und Jahr: 1991 - 1995*

Bezirk	Jahr des Verkaufs					Total
	1991	1992	1993	1994	1995	
Wedding	2 0,02	1 0,01	0 0,00	0 0,00	0 0,00	12 0,10
Prenzlauer Berg	0 0,00	1 0,01	0 0,00	0 0,00	1 0,01	7 0,06
Charlotten- burg	14 0,12	10 0,09	16 0,14	13 0,11	11 0,10	85 0,74
Spandau	112 0,97	89 0,77	129 1,12	116 1,01	104 0,90	1275 11,08
Wilmerdorf	10 0,09	18 0,16	13 0,11	18 0,16	14 0,12	150 1,30
Zehlendorf	96 0,83	103 0,90	134 1,16	107 0,93	138 1,20	1177 10,23
Steglitz	80 0,70	102 0,89	110 0,96	75 0,65	88 0,76	775 6,74
Tempelhof	124 1,08	109 0,95	146 1,27	107 0,93	124 1,08	1366 11,87
Neukölln	77 0,67	93 0,81	159 1,38	127 1,10	109 0,95	1091 9,48
Total	811 7,05	898 7,81	1262 10,97	945 8,21	942 8,19	11503 100,00

Fortsetzung auf der nächsten Seite

Tabelle B.1: *Relative und absolute Häufigkeiten der Beobachtungen nach Bezirk und Jahr: 1991 - 1995 - Fortsetzung*

Bezirk	Jahr des Verkaufs					Total
	1991	1992	1993	1994	1995	
Treptow	39	33	51	2	15	352
	0,34	0,29	0,44	0,02	0,13	3,06
Köpenick	13	34	53	42	35	454
	0,11	0,30	0,46	0,37	0,30	3,95
Lichtenberg	1	11	32	24	23	234
	0,01	0,10	0,28	0,21	0,20	2,03
Weißensee	3	0	1	1	5	259
	0,03	0,00	0,01	0,01	0,04	2,25
Pankow	16	29	47	34	30	740
	0,14	0,25	0,41	0,30	0,26	6,43
Reinickendorf	156	176	232	175	172	2227
	1,36	1,53	2,02	1,52	1,50	19,36
Marzahn	26	31	55	31	31	382
	0,23	0,27	0,48	0,27	0,27	3,32
Hohenschönhausen	7	1	23	26	9	337
	0,06	0,01	0,20	0,23	0,08	2,93
Hellersdorf	35	57	61	47	33	580
	0,30	0,50	0,53	0,41	0,29	5,04
Total	811	898	1262	945	942	11503
	7,05	7,81	10,97	8,21	8,19	100,00

Tabelle B.1 *Relative und absolute Häufigkeiten der Beobachtungen nach Bezirk und Jahr: 1996 - 2000*

Bezirk	Jahr des Verkaufs					Total
	1996	1997	1998	1999	2000	
Wedding	1	2	6	0	0	12
	0,01	0,01	0,05	0,00	0,00	0,10
Prenzlauer Berg	1	1	0	3	0	7
	0,01	0,01	0,00	0,03	0,00	0,06
Charlottenburg	5	3	3	9	1	85
	0,01	0,03	0,03	0,08	0,01	0,74
Total	1156	1094	1219	1852	1324	11503
	10,05	9,51	10,60	16,10	11,51	100,00

Fortsetzung auf der nächsten Seite

Tabelle B.1 *Relative und absolute Häufigkeiten der Beobachtungen nach Bezirk und Jahr: 1996 - 2000 - Fortsetzung*

Bezirk	Jahr des Verkaufs					Total
	1996	1997	1998	1999	2000	
Spandau	138	117	133	176	161	1275
	1,20	1,02	1,16	1,53	1,40	11,08
Wilmerdorf	18	14	18	13	14	150
	0,16	0,12	0,16	0,11	0,12	1,30
Zehlendorf	124	111	115	145	104	1177
	1,08	0,96	1,00	1,26	0,90	10,23
Steglitz	52	87	53	118	10	775
	0,45	0,76	0,46	1,03	0,09	6,74
Tempelhof	154	111	156	183	152	1366
	1,34	0,96	1,36	1,59	1,32	11,87
Neukölln	104	84	45	159	134	1091
	0,90	0,73	0,39	1,38	1,16	9,48
Treptow	30	48	48	64	22	352
	0,26	0,43	0,42	0,56	0,19	3,06
Köpenick	40	49	30	102	56	454
	0,35	0,43	0,26	0,89	0,49	3,95
Lichtenberg	25	28	27	34	29	234
	0,22	0,24	0,23	0,30	0,25	2,03
Weißensee	28	44	68	74	35	259
	0,24	0,38	0,59	0,64	0,30	2,25
Pankow	40	25	81	216	222	740
	0,35	0,22	0,70	1,88	1,93	6,43
Reinickendorf	268	225	251	339	233	2227
	2,33	1,96	2,18	2,95	2,03	19,36
Marzahn	35	45	39	46	43	382
	0,30	0,39	0,34	0,40	0,37	3,32
Hohenschönhausen	34	17	78	98	44	337
	0,30	0,15	0,68	0,85	0,38	2,93
Hellersdorf	59	83	68	73	64	580
	0,51	0,72	0,59	0,63	0,56	5,04
Total	1156	1094	1219	1852	1324	11503
	10,05	9,51	10,60	16,10	11,51	100,00

Literaturverzeichnis

- Bajari, P. und Kahn, M. (2002). Estimating housing demand with an application to explaining racial segregation in cities.
- Bayer, P. und McMillan, R. (2001). The causes and consequences of residential segregation: an equilibrium analysis of neighborhood sorting.
- Belsley, D., Kuh, E. und Welch, R. (1980). *Regression Diagnostics. Identifying influential data and sources of collinearity*, John Wiley & Sons, New York.
- Berndt, E. (1991). *The Practice of Econometrics: Classic and Contemporary*, Addison-Wesley, Reading, Massachusetts.
- Case, C. und Mayer, C. (1996). Housing price dynamics within a metropolitan area, *Regional Science and Urban Economics* (26): 387–407.
- Court, A. (1939). Hedonic price indexes with automotive examples, *The dynamics of automobile demand* pp. 99–117.
- Cutler, D. und Glaeser, E. (1995). Are ghettos good or bad?, *NBER Working Paper Series*.
- D’Agostino, R., Belanger, A. und D’Agostino jr., R. (1990). A suggestion for using powerful and informative tests of normality, *American Statistical Association* **44**(4): 316–320.
- Deller, S. und Ottem, T. (2001). Crime and the quality of life in Wisconsin counties.
- Des Rosiers, F., Lagana, A., Theriault, M. und Beaudoin, M. (1996). Shopping centres and house values: an empirical investigation, *Journal of Property Value and Investment* **14**(4): 41–62.

- Des Rosiers, F., Theriault, M., Kestens, Y. und Villeneuve, P. (2002). Landscaping and house values: An empirical investigation, *Journal of Real Estate Research* **23**(1/2): 139–161.
- Deutsche Bundesbank (1999). Monatsbericht Januar.
- Diewert, E. (2001). Hedonic regressions: A consumer theory approach.
- Downes, T. und Zabel, J. (1997). The impact of school characteristics on house prices: Chicago 1987 - 1991.
- Gerhards, H. und Keller, H. (2002). *Lexikon Baufinanzierung*, 8. Auflage, Gabler, Wiesbaden.
- Gibbons, S. (n.d.). The costs of urban property crime.
- Gordon, R. (1990). The measurement of durable goods prices, *University of Chicago Press*. A National Bureau of Economic Research Monograph.
- Griliches, Z. (1961). Hedonic price indexes for automobiles: An econometric analysis of quality change, *The Price Statistics of the Federal Government General Series*(73): 173–196. New York: National Bureau of Economic Research.
- Griliches, Z. und Ohta, M. (1976). Automobile prices revisited: Extensions of the hedonic hypothesis, in N. Terleckyj (ed.), *Practicing Econometrics: Essays in Method and Application*, number 40 in *Household Production and Consumption Studies in Income and Wealth*, National Bureau of Economic Research, Edward Elgar Publishing, MPG Books, New York, pp. 325–390.
- Harvey, A. (1994). *Ökonometrische Analyse von Zeitreihen*, 2. Auflage, R. Oldenbourg Verlag, München, Wien.
- Haurin, D. und Brasington, D. (1996). The impact of school quality on real house prices: Interjurisdictional effects.
- Hübler, O. (1989). *Ökonometrie*, Fischer, Stuttgart, New York.
- Heuer, J. (1985). *Lehrbuch der Wohnungswirtschaft*, 2. Auflage, Fritz Knapp Verlag, Frankfurt am Main.
- <http://www.stadtentwicklung.berlin.de> (n.d.).
- Judge, G. G., Hill, R. C., Griffiths, W. E., Lütkepohl, H. und Lee, T. C. (1998). *Introduction to the Theory and Practice of Econometrics.*, 2 edn, Wiley, New York.

- Lippe, G., Esemann, J. und Tänzer, T. (1994). *Das Wissen für Bankkaufleute*, 7. Auflage, Gabler, Wiesbaden.
- Luttik, J. (2000). The value of trees, water and open space as reflected by house prices in the Netherlands, *Landscape and Urban Planning* (48): 161–167.
- Moulton, B. (1995). Interarea indexes of the cost of shelter using hedonic quality adjustment techniques, *Journal of Econometrics* (68): 181–204.
- Moulton, B. (2001). The expanding role of hedonic methods in the official statistics of the United States.
- Plötz (2002). Immobilienführer Berlin, Potsdam und Umland.
- Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition, *Journal of Political Economy* (82): 34–55.
- Rosenthal, L. (2000). The value of secondary school quality.
- Ross, F., Brachmann, R. und Holzner, P. (1997). *Ermittlung des Bauwertes von Gebäuden und des Verkehrswertes von Grundstücken*, 28. Auflage, Theodor Oppermann Verlag, Hannover.
- Sieg, H., Smith, V., Banzhaf, H. und Walsh, R. (n.d.). Using locational equilibrium models to evaluate housing price indexes, *NBER Working Paper Series*.
- Sofyan, H., Schulz, R., Werwatz, A. und Witzel, R. (2002). MD*immo Immobilienpreisprognose.
- Statistisches Bundesamt (ed.) (2000). *Statistisches Jahrbuch*, Metzler-Poeschel, Wiesbaden.
- Statistisches Landesamt Berlin (ed.) (2000). *Statistisches Jahrbuch*, Kulturbuch-Verlag GmbH, Berlin.
- Theriault, M., Kestens, Y. und Des Roisiers, F. (n.d.). The impact of mature trees on house values and on residential location choices in Quebec City.
- Walden, M. (1990). Magnet schools and the differential impact of school quality on residential property values, *Journal of Real Estate Research* 5(2): 221–230.
- Woll, A. (2000). *Allgemeine Volkswirtschaftslehre*, Vahlens Handbücher der Wirtschafts- und Sozialwissenschaften, 13. Auflage, Verlag Vahlen, München.

Hiermit erkläre ich, dass ich die vorliegende Arbeit allein und nur unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe.

Edna Freese
Berlin, 22. April 2003